

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



Atty. Dkt. No. 089367-0126

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant: Takeru NAKAZATO, et al.

Title: SYSTEM AND METHOD FOR SEARCHING INFORMATION

Appl. No.: 10/822,817

Filing Date: 04/13/2004

Examiner: Unknown

Art Unit: Unknown

**CLAIM FOR CONVENTION PRIORITY**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

The benefit of the filing date of the following prior foreign application filed in the following foreign country is hereby requested, and the right of priority provided in 35 U.S.C. § 119 is hereby claimed.

In support of this claim, filed herewith is a certified copy of said original foreign application:

Japanese Patent Application No. 2003-109054  
filed 04/14/2003.

Respectfully submitted,

Date: May 18, 2004

FOLEY & LARDNER LLP  
Customer Number: 22428  
Telephone: (202) 672-5407  
Facsimile: (202) 672-5399

By Aaron C. Chettygo  
for David A. Blumenthal  
Attorney for Applicant  
Registration No. 26,257  
Reg # 41,398

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日                      2 0 0 3 年    4 月 1 4 日  
Date of Application:

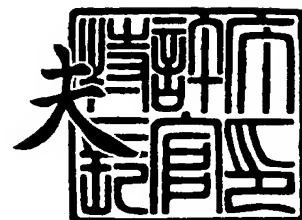
出 願 番 号                      特 願 2 0 0 3 - 1 0 9 0 5 4  
Application Number:  
[ST. 10/C] :                      [ J P 2 0 0 3 - 1 0 9 0 5 4 ]

出      願      人                      日 本 電 気 株 式 会 社  
Applicant(s):

2 0 0 4 年    3 月    2 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



【書類名】 特許願

【整理番号】 64002128

【提出日】 平成15年 4月14日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 7/22

【発明者】

【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内

【氏名】 仲里 猛留

【発明者】

【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内

【氏名】 宮川 知也

【発明者】

【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内

【氏名】 剣持 聡久

【発明者】

【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内

【氏名】 麻生川 稔

【特許出願人】

【識別番号】 000004237

【氏名又は名称】 日本電気株式会社

【代理人】

【識別番号】 100123788

【弁理士】

【氏名又は名称】 宮崎 昭夫

【電話番号】 03-3585-1882

【選任した代理人】

【識別番号】 100088328

【弁理士】

【氏名又は名称】 金田 暢之



## 【選任した代理人】

【識別番号】 100106297

【弁理士】

【氏名又は名称】 伊藤 克博

## 【選任した代理人】

【識別番号】 100106138

【弁理士】

【氏名又は名称】 石橋 政幸

## 【手数料の表示】

【予納台帳番号】 201087

【納付金額】 21,000円

## 【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0304683

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 生物情報検索システム及び検索方法

【特許請求の範囲】

【請求項 1】 ユーザから生物情報データベースへの問合せを受け取ってユーザクエリを生成して出力するユーザ入力受付部と、

前記ユーザクエリあるいは後記再検索クエリを受けて、実際に生物情報データベースを検索するための生物情報データベースクエリを生成して出力するクエリ生成部と、

前記生物情報データベースクエリに従って実際に生物情報データベースの検索を行って、検索結果を出力する生物情報データベース検索部と、

前記生物情報データベース検索部が出力する前記検索結果の解析を行って、解析結果を出力する検索結果解析部と、

前記解析結果を受けて解析結果を出力情報として出力を行うとともに再検索を行うための再検索クエリを生成して前記クエリ生成部に送る出力制御部と、

検索結果の抽出及び解析結果の出力に関する条件をあらかじめ保持しておき、これを前記検索結果解析部及び前記出力制御部に送る出力情報指定保持部とを有することを特徴とする生物情報検索システム。

【請求項 2】 さらに、検索結果の抽出及び解析結果の出力に関する条件をユーザから受け取って前記出力情報指定保持部に送る出力情報設定部を有し、前記出力情報指定保持部が、前記出力情報設定部を介して得た情報を保持する機能を有する、請求項 1 に記載の生物情報検索システム。

【請求項 3】 ユーザから生物情報データベースへの問合せ及び出力情報指定を受け取ってユーザクエリを生成し、ユーザクエリ及び出力情報指定を出力する第 2 のユーザ入力受付部と、

前記ユーザクエリに従って生物情報データベースの検索手順を設定し、生物情報データベースクエリリストとして出力する検索手順設定部と、

前記生物情報データベースクエリリスト及び前記出力情報指定に従って生物情報データベースの検索及び解析を制御し、解析完了時に結果を出力情報として出力する生物情報データベース検索制御部と、

前記生物情報データベースクエリに従って実際に生物情報データベースの検索を行って、検索結果を出力する生物情報データベース検索部と、

前記生物情報データベース検索部が出力する前記検索結果の解析を行って、解析結果を出力する検索結果解析部と、

検索結果の抽出に関する条件をあらかじめ保持しておき、これを前記検索結果解析部に送る出力情報指定保持部とを有することを特徴とする生物情報検索システム。

【請求項 4】 さらに、前記出力制御部及び前記生物情報データベース検索制御部のいずれかが出力する前記出力情報を受け取り、これを用いて文献データベースの検索を行って、文献検索結果を出力する文献データベース検索部を有する、請求項 1 から請求項 3 のいずれか 1 項に記載の生物情報検索システム。

【請求項 5】 さらに、前記文献検索結果を受けて所定の情報を抽出して文献検索解析結果を出力する文献検索結果解析部を有する、請求項 4 に記載の生物情報検索システム。

【請求項 6】 さらに、前記文献検索解析結果を受け、所定の書式に整形して出力する整形出力部を有する、請求項 5 に記載の生物情報検索システム。

【請求項 7】 さらに、前記出力制御部及び前記生物情報データベース検索制御部のいずれかが出力する前記出力情報を受け取り、遺伝子名及びタンパク質名の補完を行う検索キーワード補完部を有する、請求項 4 から請求項 6 のいずれか 1 項に記載の生物情報検索システム。

【請求項 8】 ユーザ入力受付部で、ユーザから生物情報データベースへの問合せを受け取ってユーザクエリを生成して出力し、

クエリ生成部で、前記ユーザクエリあるいは後記再検索クエリを受けて、実際に生物情報データベースを検索するための生物情報データベースクエリを生成して出力し、

生物情報データベース検索部で、前記生物情報データベースクエリに従って実際に生物情報データベースの検索を行って、検索結果を出力し、

検索結果解析部で、出力情報指定保持部が保持する検索結果の抽出条件に従って前記検索結果の解析を行って、解析結果を出力し、

出力制御部で、出力情報指定保持部が保持する解析結果の出力条件に従って、前記解析結果を受けて出力情報として出力を行うとともに、未検索分について再検索を行うための再検索クエリを生成して前記クエリ生成部に送る、各ステップを有することを特徴とする生物情報検索方法。

【請求項 9】 さらに、出力情報設定部で検索結果の抽出及び解析結果の出力に関する条件をユーザから受け取って、前記出力情報指定保持部に保持するステップを有する、請求項 8 に記載の生物情報検索方法。

【請求項 10】 前記出力制御部での前記出力情報の出力が、解析結果を受けて順次行われる、請求項 8 または請求項 9 に記載の生物情報検索方法。

【請求項 11】 前記出力制御部での前記出力情報の出力が、入力した前記解析結果を保持しておき、再検索完了時に一括して行われる、請求項 8 または請求項 9 に記載の生物情報検索方法。

【請求項 12】 ユーザ入力受付部で、ユーザから生物情報データベースへの問合せとユーザ出力情報の指定を受け取ってユーザクエリと出力指定情報とを生成して出力し、

検索手順設定部で、前記ユーザクエリに従って生物情報データベースの検索手順を設定し、生物情報データベースクエリリストおよびユーザクエリ情報として出力し、

生物情報データベース検索制御部で、前記生物情報データベースクエリリスト及び出力情報指定に従って生物情報データベースの検索を制御する生物情報データベースクエリを生成して出力し、

生物情報データベース検索部で、前記生物情報データベースクエリに従って実際に生物情報データベースの検索を行って、検索結果を出力し、

検索結果解析部で、出力情報指定保持部が保持する検索結果の抽出条件に従って前記検索結果の解析を行って、解析結果を出力し、

生物情報データベース検索制御部で、前記解析結果を受けて前記出力指定情報の指定する出力条件に従って出力情報として出力する、各ステップを有することを特徴とする生物情報検索方法。

【請求項 13】 さらに、文献データベース検索部が前記出力情報を受け取

り、これを用いて文献データベースの検索を行って、文献検索結果を出力するステップを有する、請求項 8 から請求項 12 のいずれか 1 項に記載の生物情報検索方法。

【請求項 14】 さらに、文献検索結果解析部が前記文献検索結果を受け取って、所定の情報を抽出して文献検索解析結果を出力するステップを有する、請求項 13 に記載の生物情報検索方法。

【請求項 15】 さらに、整形出力部が前記文献検索解析結果を受け取り、所定の書式に整形して出力するステップを有する、請求項 14 に記載の生物情報検索方法。

【請求項 16】 さらに、検索キーワード補完部が前記出力情報を受け取って、遺伝子名及びタンパク質名の補完を行うステップを有する、請求項 8 から請求項 15 のいずれか 1 項に記載の生物情報検索方法。

#### 【発明の詳細な説明】

##### 【0001】

##### 【発明の属する技術分野】

本発明は、生物情報データベースの検索システム及び検索方法に関し、特に別名の取得とこれを用いた文献検索の改良に関する。

##### 【0002】

##### 【従来の技術】

医学、薬学、分子生物学等の分野においては、疾患原因と発症メカニズムの解明、治療薬の開発、生物学的発見等を目的として、これまでに多くの遺伝子やタンパク質の構造や機能が解明されてきた。また、生物の発生・成長メカニズムや生体機能を細胞レベルで明らかにすべく、ヒト、マウス、ラット、線虫、イネ等様々な生物種について、ゲノム、すなわち全DNAの塩基配列を解明する研究も盛んに行われている。ゲノム解明が進むとともに、遺伝子の持つ情報に従って生成されるタンパク質を解明する研究も急速に進展している。

##### 【0003】

このような遺伝子及びタンパク質の研究を通じて、遺伝子の塩基配列、染色体上の位置、類似の遺伝子、突然変異、タンパク質を構成するアミノ酸の配列や立

体構造等、膨大なデータが得られており、各種の生物情報データベースが構築されている。

#### 【0004】

よく知られ、研究者に頻繁に利用されている生物情報データベースとして、米国の生物工学情報センター (NCBI: National Center for Biotechnology Information。以下、NCBI と略す) のEntrez (<http://www.ncbi.nlm.nih.gov/entrez/>) がある。日本の国立遺伝学研究所、ヨーロッパ分子生物学研究所 (EMBL: European Molecular Biology Laboratory。以下、EMBLと略す) の中の EBI (European Bioinformatics Institute。以下、EBIと略す) 等も、遺伝子及びタンパク質の配列の情報を中心に収めたデータベースを提供している。NCBIが提供するLocus Link (<http://www.ncbi.nlm.nih.gov/LocusLink/>)、タンパク質中心のSWISS-PROT (<http://kr.expasy.org/sprot/>) 等も、よく利用されるデータベースである。

#### 【0005】

このような状況において、医学、薬学、分子生物学等の分野の研究者は、上述のような生物情報データベースを活用することが必須となっている。例えば、疾患関連遺伝子の研究において、診断に利用できる遺伝子を特定する目的で患者と健常者の遺伝子発現を比較するような場合、予想外の遺伝子が発現し、その遺伝子に関わる詳しい情報をデータベースで調べる必要が生じることは珍しくない。新しい遺伝子を探す研究においては、DNAの断片が得られた場合、新規遺伝子かどうかは、既知のDNAに配列が一致するものがないかをデータベースで検索することによって判断が可能となる。ある遺伝子がどのような機能をもったタンパク質を生成させるかの研究においては、注目する遺伝子の配列をデータベースで調べ、配列の一部を化学合成した後、ポリメラーゼ伸張反応によって注目する遺伝子の配列を含むようなDNAを作り、タンパク質を生成させて、相互作用や電気的特性を測る等の実験を行う。これらは生物情報データベースを利用するごく一部の例であり、研究者は、生物情報データベースから様々な情報を得ながら研究を行っている。また、新規に得た知見はデータベースに登録するのが通例であり、そのため、生物情報データベースは、毎日ないし毎週程度の頻度で更新され続けている。

## 【0006】

上述のようなデータベース検索において、キーワードとして問い合わせる語および句をクエリまたはクエリーバイエグザンプル (QBE) と呼ぶがここではクエリと称する。

## 【0007】

研究者にとっては、前述のような遺伝子あるいはタンパク質の情報だけでなく、学術論文、書籍、特許文書等、文献の情報を得ることも、研究の中で非常に大きな比重を占めている。特に、世界中で発行される学術雑誌に掲載される論文を読むことは、既存の知見や、自分の研究分野における最新動向、あるいは、実験手法を知るためにも不可欠である。

## 【0008】

PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>) は、NCBI が全世界に対して無償で提供する生物医学文献データベースから生物医学文献を検索するためのツールであり、世界中の研究者により、極めて幅広くかつ頻繁に利用されている。PubMedデータベースの中核となるMEDLINE (Medical Literature, Analysis, and Retrieval System Online) は、医学、薬学、生物学等の分野の、アメリカをはじめとする70以上の国で出版された40以上の言語に渉る4600以上の雑誌の文献を収録し、1960年代半ばから現在に至るまで、1100万件以上のデータを収録している。また、新規の文献は原則的に毎日追加更新されている。研究者は、PubMed を用い、遺伝子やタンパク質の名前、著者名等のキーワードで検索を行うことにより、関連する文献のリストを得ることができる。興味のある文献のリンクをたどることにより、雑誌記事の要約を読むことができ、全文についても、研究者の所属組織が契約してあれば、さらにリンクをたどることによって読むことが可能になっている。

## 【0009】

## 【非特許文献1】

Current Issues in Molecular Biology, 3 (2001) pp.47-55

## 【非特許文献2】

Canadian Medical Association Journal, 164 (2001) pp.1317-1319

## 【0010】

## 【発明が解決しようとする課題】

しかしながら、前述のような従来のシステムでは、次のような問題がある。第一の問題点は、検索を行う研究者の知る遺伝子あるいはタンパク質の名前だけでは、検索結果に漏れが生じることがある点である。

## 【0011】

生物情報データベースを検索する際には、キーワードとして遺伝子あるいはタンパク質の名前が多く用いられる。しかし、新たに見つかった遺伝子及びタンパク質の名前に関しては、発見した研究者が各々独自の名前を付けるため、全体として名前の統一が図られていない。また、相同性の高い遺伝子あるいはタンパク質であっても、生物種が異なると、まったく別の名前がつけられていることがある。さらに、既知の遺伝子やタンパク質について新しい知見が得られ、機能に即した名前が新たに付けられたり、類似の遺伝子あるいはタンパク質を総称するような名前が新たに付けられたりすることがある。これらのことから、同じ遺伝子あるいはタンパク質に対して複数の名前が存在し、生物情報データベースでの登録名も一様でないことから、研究者個人が知っている遺伝子あるいはタンパク質の名前だけでは検索しきれないという問題、つまり検索漏れが生じる。

## 【0012】

複数ある既存の生物情報データベースにおいて、遺伝子やタンパク質の名前の記述形式が一定でないことも、検索結果に漏れを生じさせる大きな要因となっている。例えば、酵素の名前のひとつである「Carbonic Anhydrase 1」は、その略記として、「CA 1」、「CA 1」、「CA-1」等、数通りの書き方があり、アラビア数字の「1」の代わりにローマ数字の「I」が用いられることもある。検索を行う生物情報データベースで用いられている表記を検索キーワードとして指定しなければ、やはり、検索結果に漏れが生じる。

## 【0013】

以上のような検索漏れの問題は、さらに深刻な問題を引き起こす。例えば、研究者がある遺伝子やタンパク質の研究を行っているとき、自分が知る名前で既知のデータが検索されないと、登録がなされていない、つまり研究対象が新規遺伝



子あるいは新規タンパク質であると勘違いし、実験によりその遺伝子やタンパク質の試料作成を行ってしまうため、時間、労力、費用の無駄を生じる恐れがある。このような問題を避けるために、研究者は、複数の生物情報データベースでの検索を行い、あるデータベースで遺伝子の別名を得て、さらに別名で検索を行うといった作業を強いられる。これは、研究者にとって大きな負担となっている。

#### 【0014】

第二の問題点は、いくつかの生物情報データベースを検索する際には、原則として遺伝子あるいはタンパク質の名前を用いなければならない点である。

#### 【0015】

遺伝子やタンパク質には、NCBIや日本DNAデータバンクDDBJ (DNA Data Bank of Japan。以下、DDBJ と略す)等の生物情報データベースに登録した際、データベースにおけるIDとして、アクセッション番号と呼ばれる番号が付与される。このアクセッション番号は、生物情報データベースごとに独自の体系であるため、同じ遺伝子あるいはタンパク質であってもデータベースによって異なっている。このことから、複数の生物情報データベースを検索する場合、アクセッション番号では検索できないことがある。よく利用される既存の生物情報データベースを対象として、名前やアクセッション番号の異なる体系を統合する試みも始まっているが、実用にはまだ遠く、研究論文では慣用名も多く用いられるため、困難な状況と言わざるをえない。

#### 【0016】

多くの生物情報データベースで、配列や分子量等、名前以外の生物情報で検索を行えない問題もある。例えば、アクセッション番号から文献を検索しようとする際には、従来の手法では、文献検索を行う前に、NCBIのEntrezデータベースでアクセッション番号を検索し、その遺伝子あるいはタンパク質の名前を調べておかななくてはならない。これは使用者にとって大きな負担である。また、新規の遺伝子あるいはタンパク質の候補となる試料を得た場合は、まだ名前やアクセッション番号が付けられていない可能性があるため、まずは、塩基あるいはアミノ酸配列を用いてデータベースの検索を行うのが一般的である。

#### 【0017】

従来の手法では、塩基あるいはアミノ酸配列の相同性を調べるツールであるBLAST (Basic Local Alignment Search Tool。以下、BLASTと略す) を用い、相同性のある遺伝子あるいはタンパク質を調べた後、その結果から、研究者が遺伝子やタンパク質の名前を読み取り、PubMedのような文献検索ツールにキーワードとして名前を手入力して検索を行っていた。これら一連の作業は、研究者にとって少なからぬ負担であった。

#### 【0018】

本発明の目的は、検索を行う研究者の知る遺伝子あるいはタンパク質の名前だけで検索を行っても、検索結果の漏れが生じにくい生物情報検索システム及び検索方法を提供することにある。

#### 【0019】

本発明の別の目的は、遺伝子あるいはタンパク質の名前以外を検索キーワードにしても、これらの名前以外で検索できない生物情報データベースまで検索を行うような、生物情報検索システム及び検索方法を提供することにある。

#### 【0020】

さらに、本発明の別の目的は、以上のような生物情報検索システム及び検索方法を提供することにより、研究者が生物情報データベースを研究する際の負担を軽減することにある。

#### 【0021】

##### 【課題を解決するための手段】

本発明の生物情報検索システム及び検索方法は、ユーザ入力受付部を有することから、ユーザから生物情報データベースへの問合せを受け取ってユーザクエリを生成して出力することができ、クエリ生成部を有することから、ユーザクエリあるいは後記の再検索クエリを受けて、実際に生物情報データベースを検索するための生物情報データベースクエリを生成して出力することができ、生物情報データベース検索部を有することから、生物情報データベースクエリに従って実際に生物情報データベースの検索を行って検索結果を出力することができ、検索結果解析部を有することから、生物情報データベース検索部が出力する検索結果の解析を行って、解析結果を出力することができ、出力制御部を有することから、

検索結果を受けて出力情報の出力を行うとともに再検索を行うための再検索クエリを生成してクエリ生成部に送ることができ、出力情報指定保持部を有することから、検索結果の抽出及び解析結果の出力に関する条件をあらかじめ保持しておき、これを検索結果解析部及び出力制御部に送ることができる。

#### 【0022】

さらに、本発明の生物情報検索システム及び検索方法は、出力情報設定部を有することから、検索結果の抽出及び解析結果の出力に関する条件をユーザから受け取って出力情報指定保持部に送ることができ、出力情報指定保持部が、出力情報設定部を介して得た情報を保持することができる。

#### 【0023】

さらに、本発明の生物情報検索システム及び検索方法は、出力制御部が、解析結果を順次出力情報として出力することもでき、あるいは解析結果を保持しておくことにより、再検索完了時に一括して出力情報の出力を行うことができる。

#### 【0024】

また、本発明の生物情報検索システム及び検索方法は、生物情報データベース検索部、検索結果解析部、出力情報指定保持部に加えて、第2のユーザ入力受付部を有することから、ユーザから生物情報データベースへの問合せ及び出力情報指定を受け取ってユーザクエリを生成し、ユーザクエリ及び出力情報指定を出力することができ、検索手順設定部を有することから、ユーザクエリに従って生物情報データベースの検索手順を設定し、生物情報データベースクエリリストとして出力することができ、生物情報データベース検索制御部を有することから、生物情報データベースクエリリスト及び出力情報指定に従って生物情報データベースの検索を制御し、検索・解析完了時に結果を出力情報として出力することができる。

#### 【0025】

さらに、本発明の生物情報検索システム及び検索方法は、上記の手段に加えて、文献データベース検索部を有することにより、出力制御部あるいは生物情報データベース検索制御部が出力する出力情報を受け取り、これを用いて文献データベースの検索を行って、文献検索結果を出力することができる。

**【0026】**

さらに、本発明の生物情報検索システム及び検索方法は、上記の手段に加えて、文献検索結果解析部を有することにより、文献検索結果を受けて所定の情報を抽出して文献検索解析結果を出力することができる。

**【0027】**

さらに、本発明の検索システム及び検索方法は、上記の手段に加えて、整形出力部を有することにより、文献検索解析結果を受け、所定の書式に整形して出力することができる。

**【0028】**

さらに、本発明の生物情報検索システム及び検索方法は、上記の手段に加えて、文献検索キーワード補完部を有することにより、出力制御部及び生物情報データベース検索制御部が出力する出力情報を受け取り、遺伝子及びタンパク質の名前の補完を行うことができる。

**【0029】****【発明の実施の形態】**

本発明の実施の形態について、図面を参照して詳細に説明する。図1のブロック構成図を参照すると、本発明の第一の実施形態としての生物情報検索システム及び検索方法が示されている。

**【0030】**

図1に示すように、本発明の生物情報検索システム及び検索方法は、ユーザ入力受付部11と、クエリ生成部12と、生物情報データベース検索部13と、検索結果解析部14と、出力制御部15と、出力情報指定保持部16からなる。

**【0031】**

ユーザ入力受付部11は、ユーザ入力101を受け取る。ユーザ入力101は、例えば、遺伝子あるいはタンパク質の名前、アクセッション番号、塩基あるいはアミノ酸配列等の、生物情報データベースへの検索キーワードを含んでいる。

**【0032】**

ユーザ入力受付部11は、例えば、ユーザ入力101の種類に応じた入力欄を備えることにより、ユーザ入力101が含んでいる検索キーワードそれぞれに種

類を対応付ける。あるいは、ユーザ入力受付部 11 は、入力の種類に応じた書式・形式を保持しており、アクセッション番号の形式、あるいは配列とその要素の記述形式等に基づいて、ユーザ入力 101 が含む検索キーワードに種類を対応付ける。ここで、検索キーワードの種類を 1 つに特定できない入力、ユーザ入力 101 がとりうるすべての種類に対応付ける。

#### 【0033】

以上のように、ユーザ入力部 11 は、ユーザ入力 101 が含む検索キーワードそれぞれに種類を対応付けた後、検索キーワードとその種類の組のリストとして、ユーザクエリ 111 をクエリ生成部 12 に送る。

#### 【0034】

クエリ生成部 12 は、ユーザ入力受付部 11 から、ユーザクエリ 111 を受け取る。

#### 【0035】

クエリ生成部 12 は、各種データベースの所在と問い合わせ可能な情報の種類を組にして、データベース情報として保持している。クエリ生成部 12 は、さらに、データベース情報に対応する検索済みフラグのテーブルをもっている。クエリ生成部 12 は、ユーザクエリ 111 を受け取ると、すべての検索済みフラグを未検索の状態に初期化する。続いて、ユーザクエリ 111 が含む情報の種類とデータベース情報が含む情報の種類を比較し、問合せを行う生物情報データベースの所在を決定する。ここで、データベースの所在は、例えば URL (Uniform Resource Locator。以下、URL と略す) で表すことができる。

#### 【0036】

これら、生物情報データベースの所在、検索キーワードとその種類から、クエリ生成部 12 は、実際にデータベース検索を行うための生物情報データベースクエリ 121 を生成し、生物情報データベース検索部 13 に送る。このとき、クエリ生成部 12 は、生物情報データベースクエリ 121 で検索を行うデータベースとデータの種類について、検索済みフラグを検索済みに変更する。

#### 【0037】

生物情報データベース検索部 13 は、クエリ生成部 12 から、生物情報データ

ベースクエリ 121 を受け取ると、これを問合せ 131 として指定された生物情報データベース 999 に送信し、生物情報データベース 999 から検索結果 132 を受け取ると、その結果を生物情報データベース検索結果 133 として検索結果解析部 14 に送る。

#### 【0038】

生物情報データベース 999 のサーバのエラーが発生した場合は、検索結果 132 の内容からエラーであることがわかり、また、所定時間内に検索結果 132 が得られなければ、通信障害が発生したことがわかる。これらのエラーが生じた場合、生物情報データベース検索部 13 は、クエリ生成部 12 にエラー情報 134 を送る。もちろん、所定回数の再検索を行っても検索結果 132 が得られない時にエラー情報 134 を送ってもよく、これには再検索を行う回数をあらかじめ与え、エラー発生時に再検索を繰り返すようにしておけばよい。クエリ生成部 12 はエラー情報 134 を受け取ると、検索済みフラグを未検索の状態に戻す。

#### 【0039】

検索結果解析部 14 は、生物情報データベース検索部 13 から、生物情報データベース検索結果 133 を受け取ると、出力情報指定保持部 16 に抽出条件要求 141 を送る。抽出条件要求 141 は、生物情報データベース検索結果 133 に基づいて、例えば、どの生物情報データベースの検索を行ったか、どの種類のデータを検索したか等の情報を含む。出力情報指定保持部 16 は抽出条件要求 141 の内容に応じた抽出条件 161 を検索結果解析部 14 に返す。抽出条件 161 は、各データベースの検索結果について、出力として用いる部分を特定するための情報や抽出のルール、検索しても結果が得られない場合に各生物情報データベースが返す情報の書式等を含む。検索結果解析部 14 は、抽出条件 161 の内容に従って、生物情報データベース検索結果 133 から、データベースの種類、検索キーワードとその種類のリストを抽出し、解析結果 142 として、出力制御部 15 に送る。

#### 【0040】

出力制御部 15 は、検索のはじめに出力情報指定保持部 16 から出力情報指定 162 を受け取り、これを出力フラグセットとして保持する。出力情報指定 16

2 は、データベースとその出力の種類の各組に対応する出力フラグの初期値を含む。

#### 【0041】

出力制御部 15 は、検索結果解析部 14 から解析結果 142 を受け取ると、これを保持するとともに、解析結果 142 が出力情報指定 162 で指定するデータベースと情報の種類の組を含むかどうかを調べる。解析結果 142 が出力情報指定 162 で指定する内容を含むならば、出力フラグセット中で該当するフラグを、結果取得済みを示す値に変更する。

#### 【0042】

出力情報指定 162 で指定された結果がすべて得られたとき、すなわち、出力フラグセットの全フラグがすべて結果取得済みとなったとき、出力制御部 15 は、出力 152 として保持しておいたすべての解析結果 142 を一括して出力する。出力すべき内容が全部得られていないとき、つまり、出力フラグセットの中に結果取得済みでないものがある場合には、保持しているすべての解析結果 142 から検索キーワードとその種類の情報を組にして抽出し、その組のリストを、再検索クエリ 151 としてクエリ生成部 12 に送る。

#### 【0043】

出力制御部 15 の別の形態として、解析結果 142 を逐次出力するようにしてもよい。それには、受け取った解析結果 142 を、受け取るたびに出力 152 として出力しつつ、上述のような出力フラグセットの変更も行う。出力すべき内容が全部得られていないとき、つまり、出力フラグの中に結果取得済みでないものがある場合には、保持しているすべての解析結果 142 から検索キーワードとその種類を組にして抽出し、その組のリストを、再検索クエリ 151 としてクエリ生成部 12 に送る。出力すべき内容が全部得られたとき、つまり、出力フラグセットの全フラグが検索済みとなったとき、検索を完了する。このとき検索完了を示す情報を出力すれば、利用者に検索完了を明示することが可能になる。

#### 【0044】

再検索クエリ 151 を受け取った後のクエリ生成部 12 での処理は、再検索クエリ 151 の内容がユーザクエリ 111 の内容に置き換わったとすれば前述と同

様であり、それ以降の生物情報データベース検索部 13、検索結果解析部 14、出力制御部 15 の各処理も同様である。ただし、再検索クエリ 151 を受け取ったとき、クエリ生成部 12 は、検索済みフラグの初期化は行わず、検索済みフラグが検索済みの情報の組に対しては、生物情報データベースクエリ 121 の内容を、検索済みであることを示す情報に変更する。また、再検索クエリ 151 が含むすべての組で検索が既に行われていた場合、クエリ生成部 12 は、検索完了情報 122 を出力制御部 15 に送る。出力制御部 15 は、検索完了情報 122 を受け取ると、出力フラグセットの状態に関わらず、出力 152 として保持しておいたすべての解析結果 142 を一括して出力する。あるいは、検索を完了する。もちろん、このとき検索完了を示す情報を出力してもよい。

#### 【0045】

出力情報指定保持部 16 は、前述のような抽出条件 161 をあらかじめ保持しており、検索結果解析部 14 から抽出条件要求 141 を受け取ると、これを検索結果解析部 14 に抽出条件 161 として送る。

#### 【0046】

また、出力情報指定保持部 16 は、前述のような出力情報指定 162 をあらかじめ保持しており、出力制御部 15 がユーザクエリ 111 に基づく最初の解析結果 142 を受け取るとき、出力情報指定保持部 16 は、出力制御部 15 に出力情報指定 162 を送る。

#### 【0047】

図 2 のフローチャートを参照すると、ユーザ入力受付部 11 の処理の詳細が示されている。ユーザ入力受付部 11 はユーザ入力 101 を受け取ると、まず、ユーザ入力 101 が含む検索キーワードに対して、その種類を対応付ける。ユーザ入力受付部 11 は、前述のような入力の種類に応じた入力欄を持っている。この欄の違いから、ユーザ入力 101 が含んでいる検索キーワードそれぞれに種類を対応付ける。

#### 【0048】

ユーザ入力受付部 11 の別の形態として、ユーザ入力受付部 11 が、入力の種類に応じて、公に提供され、よく知られたデータの書式・形式を保持しておくよ



うにしてもよい。この場合、ユーザ入力受付部 1 1 は、まず、ステップ 1 1 0 1 で、ユーザ入力 1 0 1 が含む検索キーワードが保持している書式・形式にのっとった記述でないかを確認する。例えば、FASTA形式は、同業者の中ではよく知られた形式で、一行目に「>」に続いて遺伝子あるいはタンパク質の名前が、二行目以降にその塩基あるいはアミノ酸配列が続く書式である。また、遺伝子名、タンパク質名、塩基配列、アミノ酸配列、生物種、発表学術論文といった情報が定型化されているGenBank形式も同業者の中ではよく知られた形式である。これらの書式にのっとっていた場合、ステップ 1 1 0 2 において、ユーザ入力受付部 1 1 があらかじめ保持している形式に基づいて情報を切り出し、データの種類の対応付けを行う。のっとっていない場合、ステップ 1 1 0 3 に進む。

#### 【0 0 4 9】

ステップ 1 1 0 3 は、ユーザ入力 1 0 1 の文字数を算出し、所定数以上（たとえば 2 0 字以上）であれば、ユーザ入力 1 0 1 を配列であるに対応付ける。配列であると判定した場合には、ステップ 1 1 0 5 において、配列の種類を対応付ける。配列情報に含まれる文字が、G, A, T, Cのみ（もしくはそれに加えて、X, N等）しか含まない場合は、塩基配列、それ以外の文字を含む場合は、アミノ酸配列と対応付けることができる。

#### 【0 0 5 0】

配列でないとみなした場合、ステップ 1 1 0 4 に進む。ステップ 1 1 0 4 では、ユーザ入力 1 0 1 が、アクセッション番号の所定の書式（英数字の組み合わせ）にのっとっているかを判定する。この書式・形式の情報は、ユーザ入力受付部 1 1 があらかじめ保持しており、それとユーザ入力 1 0 1 を比較する。データベースの種類としては、例えばEntrez、LocusLink、SWISS-PROT 等が挙げられる。

#### 【0 0 5 1】

仮に、検索キーワードの種類を 1 つに特定できない入力は、ユーザ入力 1 0 1 がとりうるすべての種類に対応付ける。

#### 【0 0 5 2】

図 3 のフローチャートを参照すると、クエリ生成部 1 2 の処理の詳細が示されている。クエリ生成部 1 2 は、ユーザクエリ 1 1 1 あるいは再検索クエリ 1 5 1

として、検索キーワードとその種類の組のリストを受け取る。

#### 【0053】

ステップ1201では、すでに検索済みフラグのテーブルがつくられているかを判定する。検索済みフラグのテーブルがない場合、それは、ユーザ入力受付部11からユーザクエリ111を受け取ったことを表し、次にステップ1202に進む。検索済みフラグのテーブルがすでにつくられている場合、それは、出力制御部15から再検索クエリ151を受け取ったことを表し、続いてステップ1205に進む。

#### 【0054】

ステップ1202では、データベース情報として保持しているデータベースの所在と問い合わせ可能な情報の種類の組を読み込む。続いて、ステップ1203では、保持しているデータベースの所在及び情報の種類の各組に対応する検索済みフラグのテーブルを読み込む。ステップ1204では、すべての検索済みフラグを未検索の状態に初期化する。

#### 【0055】

ステップ1205では、ユーザクエリ111あるいは再検索クエリ151における種類の情報と、データベース情報が含む種類の情報を比較し、問合せを行う生物情報データベースの所在を決定する。ここで、データベースの所在は、例えばURLで表すことができる。

#### 【0056】

これら、生物情報データベース、検索キーワードとその種類から、クエリ生成部12は、実際にデータベース検索を行う生物情報データベース検索クエリ121を生成する。

#### 【0057】

仮に、問い合わせのできるデータベースが見つからなかった場合は、検索不能の情報をその組について付与する。

#### 【0058】

ステップ1206では、ステップ1205で決定した生物情報データベースクエリ121に対して、その組で既に検索がなされていないかを判定する。もしも

既に検索がなされていた場合、ステップ1207において、そのデータベースとクエリの種類の組に対して検索済みフラグが立てられている。この検索済みフラグの状態によって、既に検索が行われたか否かを判定する。もしも検索が行われていなかった場合、ステップ1207において、生物情報データベースクエリ121に対応する検索済みフラグを検索済みに変更する。

#### 【0059】

もしも検索が行われていた場合、ステップ1208において、クエリ生成部12で決定した生物情報データベースクエリ121について、その内容を検索済みであることを示す情報に書き換え、ステップ1209において、全部のクエリが検索済みか判定する。

#### 【0060】

もし、入力されたリストのすべての組について、それが検索済みだった場合、ステップ1210において、検索完了情報122を出力制御部15に送る。

#### 【0061】

図4のフローチャートを参照すると、生物情報データベース検索部13の処理の詳細が示されている。生物情報データベース検索部13は、生物情報データベースクエリ121を入力とする。

#### 【0062】

ステップ1301では、指定された生物情報データベース999に生物情報データベースクエリ121を問合せ131として送信する。すると、実際の検索が始まり、ステップ1302では、検索結果が出るのを待つ。

#### 【0063】

仮に、通信状態、あるいは実際の検索を行うサーバの問題により、所定時間内に結果が得られなかった場合、通信障害が発生したとみなし、ステップ1302において、タイムアウトとみなし、ステップ1304でクエリ生成部12にエラー情報134を送る。クエリ生成部12はエラー情報134を受け取ると、検索済みフラグセットを未検索の状態に戻す。

#### 【0064】

もちろん、所定回数の再検索を行っても検索結果132が得られないときに、

エラー情報 134 を送ってもよく、これには、エラー発生時にあらかじめ与えた上限回数まで検索を繰り返すようにしておけばよい。この場合、図 5 のフローチャートに示すように、ステップ 1302 においてタイムアウトとなった後、ステップ 1305 において、再検索カウンタを増やす処理を行う。次に、ステップ 1306 において再検索カウンタの値が、所定の回数以上かどうか判定する。もし所定回数以上だった場合、クエリ生成部 12 にエラー情報 134 を送る。規定回数に満たない場合は、ステップ 1301 に戻り、再び検索を試みる。

#### 【0065】

一方、所定時間内に結果が返ってきた場合、ステップ 1303 では、検索を行っているデータベースから検索結果 132 を得る。その結果を、生物情報データベース検索結果 133 として、検索結果解析部 14 に送る。

#### 【0066】

図 6 のフローチャートを参照すると、検索結果解析部 14 の処理の詳細が示されている。

#### 【0067】

ステップ 1401 では、生物情報データベース検索部 13 から、生物情報データベース検索結果 133 を受け取る。

#### 【0068】

ステップ 1402 では、出力情報指定保持部 16 に抽出条件要求 141 を送る。ステップ 1403 では、出力情報指定保持部 16 から、抽出条件 161 を受け取る。

#### 【0069】

ステップ 1404 では、生物情報データベース検索結果 133 より、抽出条件 161 に従って、データベースの種類、検索キーワードとその種類のリストを抽出する。

#### 【0070】

このとき、例えば、遺伝子あるいはタンパク質の名前からアクセッション番号を検索した際には、生物種、突然変異等が異なるデータが複数登録されているため、名前に対応するアクセッション番号が複数得られる。このように抽出可能な

データが複数存在し、そのうちひとつについて情報を得ればよいというとき、名前の付け方、アクセス番号の付け方、出てくる順番等、所定の判定規則に従って、データの内容としてひとつを保持する。あるいは、抽出可能なデータをすべて抽出し、それらを生物情報データベースの検索キーワードとして複数キーワードをOR条件で結合して用いる形式で保持してもよい。

#### 【0071】

ステップ1405では、抽出したデータベースの種類、検索キーワードとその種類のリストを解析結果142として出力制御部15に送る。

#### 【0072】

図7のフローチャートを参照すると、出力制御部15の第一の形態の処理の詳細が示されている。出力制御部15の第一の形態は、逐次得られる解析結果142を保持しておき、検索完了時に一括出力を行う形態である。

#### 【0073】

ステップ1501では、検索結果解析部14から、解析結果142を受け取る。

#### 【0074】

ステップ1502では、すでに出力フラグセットが設定されているかを確認する。再検索が行われて、その解析結果を出力制御部15が受け取った場合、すでに出力フラグセットが設定されている。出力フラグセットが設定されている場合、ステップ1506に進む。設定されていない場合、ステップ1503に進む。

#### 【0075】

ステップ1503では、出力情報指定保持部16から、出力情報指定162を受け取る。ステップ1504では、受け取った出力情報指定162を、出力フラグセットとして保持する。出力には、この出力フラグセットで設定された条件を満たすことが必要である。そして、ステップ1505では、すべての出力フラグを結果未取得の状態に初期化する。

#### 【0076】

次に、ステップ1506では、受け取った解析結果142を保持しておく。ステップ1507では、出力フラグセットのデータベースとデータの種類の組と、

解析結果 142 に含まれるデータベースの種類とデータの種類の組との比較を行う。そして、出力フラグに対応するものが解析結果 142 に含まれるとき、そのフラグを結果取得済みにかえる。

#### 【0077】

ステップ 1508 では、出力フラグセットの状態により、出力するか、再検索を行うかを判定する。出力フラグセットの全フラグがすべて結果取得済みとなったとき、ステップ 1509 で、保持しておいたすべての解析結果 142 を出力 152 として一括して出力する。出力すべき内容が全部得られていないとき、つまり、出力フラグセットの中に結果取得済みでないものがある場合には、ステップ 1510 で、保持しているすべての解析結果 142 から検索キーワードとその種類の情報を組にして抽出し、ステップ 1511 で、その組のリストを再検索クエリ 151 としてクエリ生成部 12 に送る。

#### 【0078】

図 8 のフローチャートを参照すると、出力制御部 15 の別の形態の処理の詳細が示されている。出力制御部 15 の別の形態は、解析結果 142 を得るたびに逐次出力を行う形態である。

#### 【0079】

ここで、出力制御部 15 の別の形態の詳細において、ステップ 1512、ステップ 1513 以外は出力制御部 15 の第一の形態の詳細と同じである。よって、以下では、主にステップ 1512、ステップ 1513 の動作について説明する。

#### 【0080】

出力制御部 15 の別の形態では、検索結果解析部 14 から、解析結果 142 を受け取り、出力フラグセットの設定を行った後、ステップ 1512 で、受け取った解析結果 142 を出力 152 として出力する。そして、解析結果に対応する出力フラグを結果取得済みにした後、出力フラグセットの状態により、再検索を行うか、検索を終了するかを判定する。再検索を行う際の処理は前述の 1510 および 1511 と同様である。

#### 【0081】

検索を終了する場合は、ステップ 1513 において、検索を終了する。

**【0082】**

ここで、図9のフローチャートに示すように、ステップ1514として、検索を終了する時に検索完了を示す情報を出力してもよい。これにより、利用者に検索完了を明示することが可能になる。

**【0083】**

図10のフローチャートを参照すると、出力情報指定保持部16の処理の詳細が示されている。ステップ1601では、検索結果解析部14から、抽出条件要求141を受け取る。

**【0084】**

ステップ1602では、抽出条件要求141に従って、生物情報データベース検索結果133から出力させる部分を特定するための情報や抽出のルールを抽出条件として読み込む。ステップ1603では、読み込んだ抽出条件161を、検索結果解析部14に送る。

**【0085】**

ステップ1604では、データベースとその出力の種類の各組に対応する出力フラグの初期値を出力情報指定162として読み込む。ステップ1605では、読み込んだ出力情報指定162を出力制御部15に送る。

**【0086】**

次に、第一の実施形態の具体的な実施例について、図面を参照して詳細に説明する。以下は、ユーザが遺伝子あるいはタンパク質の名前の検索キーワードとして「pendrin」と入力し、EntrezとLocusLinkからはその別名を取得し、また、Entrezで別名検索を行う過程で得た配列情報を用いてBLASTで相同性検索を行って、その結果から相同性の高い遺伝子あるいはタンパク質の名前を取得する場合を示している。

**【0087】**

図11は、ユーザによるキーワードの入力画面を示している。図11に示すように、本実施例では、ユーザによって、遺伝子あるいはタンパク質の名前、アクセスセッション番号、配列等の指定が行える。

**【0088】**

ユーザが「Protein or gene name」欄に「pendrin」と入力し、Submit ボタンをクリックすると、ユーザ入力受付部 11 は、「pendrin」の種類は名前である  
と対応付ける。入力欄によって対応付けを行う他にも、第一の実施形態の説明で述べたように、アクセッション番号の書式である英字と数字の組み合わせではないこと、及び、塩基あるいはアミノ酸の配列のように決まった文字から成るある程度の長さの情報ではないことから判別は可能である。

#### 【0089】

そして、ユーザ入力受付部 11 は、ユーザクエリ 111 として、検索キーワードが「pendrin」であり、その種類が遺伝子あるいはタンパク質の名前であるという情報をクエリ生成部 12 に送る。

#### 【0090】

クエリ生成部 12 は、上記のようなユーザクエリ 111 を受け取ると、データベース情報として保持しているデータベースの所在と問い合わせ可能な情報の種類の組を読み込み、さらに、検索済みフラグのテーブルを読み込んですべてのフラグを初期化する。本実施例では、データベース情報は、{データベースの所在、受け付ける入力の種類} として、次のような情報を持つものとする。

Entrez 1 = {Entrezの所在 1、遺伝子あるいはタンパク質の名前}

Entrez 2 = {Entrezの所在 2、GI番号}

LocusLink 1 = {LocusLinkの所在 1、遺伝子あるいはタンパク質の名前}

BLAST 1 = {BLASTの所在 1、配列情報}

ここで、「Entrezの所在 1」とは、EntrezそのもののURLにCGI (Common Gateway Interface) による検索文字列のテンプレートを結合したものとする。テンプレートとは、検索キーワード部分のみが置換可能な変数名に置き換えられたものである。「Entrezの所在 2」、「LocusLinkの所在 1」、「BLASTの所在 1」等も同様のものとする。もちろん、URLはローカルなデータベースの所在も指定できる。

#### 【0091】

このようなデータベース情報に基づき、検索済みフラグのテーブルは、その要素である {データベース、検索済みフラグ} として、次のような情報を持つ。



Entrez1フラグ = {Entrez1、FALSE}

Entrez2フラグ = {Entrez2、FALSE}

LocusLink1フラグ = {LocusLink1、FALSE}

BLAST1フラグ = {BLAST1、FALSE}

ここで、各フラグがFALSEであるのは、初期化によって、未検索と設定されていることを示す。

#### 【0092】

ユーザクエリ111によると、ユーザの入力によって、遺伝子あるいはタンパク質の名前であるpendrinが与えられているので、上記のようなデータベース情報から、クエリ生成部12は、Entrezの所在1及びLocusLinkの所在1を用い、検索キーワードとしてpendrinを用いて検索を行うことを決定する。そして、Entrez1フラグ及びLocusLink1フラグを、各々、{Entrez1、TRUE}、{LocusLink1、TRUE}のように変更するとともに、Entrezの所在1及びLocusLinkの所在1に基づいて生物情報データベースクエリ121を生成して生物情報データベース検索部13に送る。ここで生成された生物情報データベースクエリ121は、本実施例では、Entrez及びLocusLinkそのもののURLにCGIによる検索文字列を結合したものであり、検索文字列のテンプレート中の検索キーワードがpendrinに置き換えられたものである。

#### 【0093】

生物情報データベース検索部13は、生物情報データベースクエリ121の内容に従ってEntrez及びLocusLinkに対してクエリを送信し、これらのデータベースから検索結果132が得られたら、どの生物情報データベースにどのような種類のキーワードで検索を行ったかの情報を検索結果132に付加し、これを生物情報データベース検索結果133として検索結果解析部14に送る。

#### 【0094】

検索結果解析部14は、生物情報データベース検索結果133を受け取ると、出力情報指定保持部16に抽出条件要求141を送る。抽出条件要求141は、ここでは、検索を行った生物情報データベースがEntrez及びLocusLinkであること、各々のデータベースを遺伝子あるいはタンパク質の名前で検索したことを含

む。抽出条件要求 141 に応じて、出力情報指定保持部 16 は抽出条件 161 を返し、検索結果解析部 14 はこれを受け取る。抽出条件 161 には、Entrez 及び LocusLink を遺伝子あるいはタンパク質の名前で検索したときに、生物情報データベース検索結果 133 がどのような書式であるか、及び、生物情報データベース検索結果 133 のどの部分を抽出するかが記述されている。ここでは、抽出条件 161 には、LocusLink からは別名を抽出すること、Entrez からは遺伝子あるいはタンパク質の名前、アクセッション番号、GI 番号を抽出することが記述されているとする。

#### 【0095】

通常の Web ブラウザを用いた検索において、LocusLink で遺伝子あるいはタンパク質の名前として pendrin を検索すると、図 12 に示すような結果が得られる。図 12 に示すように、その結果は、所定の場所に、略号、略さない形式の名前、LocusLink、冗長性を排除した遺伝子あるいはタンパク質のデータベースである RefSeq や疾患に関連する遺伝子あるいはタンパク質のデータベースである OMIM 等の生物情報データベースにおけるアクセッション番号、遺伝子の染色体上の位置等が表示される。Symbol、Description 及び Aliases の欄には、遺伝子及びタンパク質の名前に関する情報が書かれている。これらのことから、生物情報データベース検索の結果として、通常の Web ブラウザによる検索と同様の結果が得られるならば、HTML (Hyper Text Markup Language) 形式の検索結果中、所定位置もしくは書式の部分を抽出すれば、遺伝子やタンパク質の別名、アクセッション番号等が得られることがわかる。

#### 【0096】

本実施例においては、抽出条件 161 が、LocusLink の検索結果については Symbol、Description 及び Aliases の欄に記載された内容を抽出することを指定しておくことにより、pendrin の別名として、「PDS」、「SLC26A4」、「DFNB4」等を抽出する。

#### 【0097】

また、通常の Web ブラウザを用いた検索において、Entrez で遺伝子あるいはタンパク質の名前として pendrin を検索すると、図 13 に示すような結果が得られ

る。図13に示すように、その結果は、所定の場所に、GenBankあるいはRefSeqのアクセッション番号、GI番号、遺伝子あるいはタンパク質の名前等が表示される。LocusLinkの場合と同様に、HTML形式で得られる検索結果中、所定位置もしくは書式の部分を抽出すれば、遺伝子やタンパク質の名前、アクセッション番号、GI番号等の情報が得られる。

#### 【0098】

本実施例においては、抽出条件161として、「名前 mRNA, complete cds」と記述されている部分と、RefSeqに基づくアクセッション番号が記述されている「|ref|NW#03954.8|」の部分抽出することを指定しておく。また、ヒトのデータを得るよう、名前には「Homo sapiens」を含むことも条件としておく。これにより、名前として「Homo sapiens pendrin (PDS) mRNA, complete cds」が、アクセッション番号としてAF030880が、GI番号として2654004が得られる。ここではヒトのデータを抽出したが、もちろん、他の生物種を指定することも容易である。

#### 【0099】

検索結果解析部14は、LocusLinkから得た別名、Entrezから得た名前、アクセッション番号、GI番号等を解析結果142として出力制御部15に送る。

#### 【0100】

出力制御部15は、解析結果142を受け取ってこれを保持するとともに、出力フラグの変更を行う。本実施例は、LocusLinkとEntrezから別名を取得し、BLASTから名前を取得するものである。出力フラグとしては、LocusLinkによる別名のフラグ、Entrezによる別名のフラグ、BLASTによる名前のフラグが設定されている。受け取った解析結果142には、LocusLinkの別名が含まれているので、LocusLinkによる別名のフラグのみ結果取得済みに変更する。

#### 【0101】

次に出力制御部15は、出力フラグがすべて結果取得済みかどうかを調べる。ここでは、Entrezによる別名のフラグ、BLASTによる名前のフラグが結果取得済みになっていない。よって、解析結果142から、名前としてpendrin、SLC26A4、PDS及びDFNB4を、アクセッション番号としてAF03880を、GI番号として265004

を抽出し、これらの種類とキーワードの組を再検索クエリ 151 としてクエリ生成部 12 に送る。

#### 【0102】

クエリ生成部 12 は、出力制御部 15 から再検索クエリ 151 を受け取ると、検索済みフラグと照らし合わせて、検索済みの組み合わせが無いかを調べる。LocusLink フラグ及び Entrez 1 フラグから、LocusLink と名前、Entrez と名前の組み合わせは既に検索済みであることがわかる。次にクエリ生成部 12 は、データベース情報を参照し、再検索クエリ 151 が含む組み合わせで検索可能なデータベースを調べる。すると、Entrez 2 から、Entrez と GI 番号の組み合わせが検索可能であることがわかるので、前述の 1 回目の検索と同様にして、Entrez 2 フラグを検索済みに変更するとともに、生物情報データベースクエリ 121 を生成する。この 2 回目の生物情報データベースクエリは、GI 番号 265004 で Entrez を検索することを示すものとなる。

#### 【0103】

生物情報データベース検索部 13 は 1 回目の検索と同様にして検索を行って、生物情報データベース検索結果 133 として、GI 番号 265004 で Entrez を検索した結果を検索結果解析部 14 に送る。

#### 【0104】

通常の Web ブラウザを用いて GI 番号 265004 で Entrez を検索すると、図 14 に示すような画面を得ることができる。図 14 に示すように、その結果には、遺伝子名、タンパク質名、生物種、文献情報、塩基配列、アミノ酸配列等の情報が含まれている。ここで、gene 及び product の欄には、遺伝子あるいはタンパク質の名前に関する情報が記述されており、pendrin 以外は別名とみなすことができる。また、ORIGIN の欄には塩基配列が、translation の欄にはアミノ酸配列が記されている。

#### 【0105】

Entrez を GI 番号で検索した結果について、本実施例の抽出条件 161 は、gene 及び product の欄から別名を抽出し、ORIGIN の欄から塩基配列を抽出し、translation の欄からアミノ酸配列を抽出することを記述しておくものとする。これによ

り、EntrezをGI番号で検索した結果から、別名として「PDS」を抽出することができる。また、ORIGIN及びtranslationの欄から、それぞれ塩基配列及びアミノ酸配列を抽出することができる。配列情報は長いため、以下の説明でも省略する。

#### 【0106】

検索結果解析部14は、このようにして得たEntrezからの別名、塩基配列、アミノ酸配列を解析結果142として出力制御部15に送る。

#### 【0107】

出力制御部15は、解析結果142を受け取ってこれを保持するとともに、出力フラグの変更を行う。ここで得た解析結果142は、Entrezから得た別名を含むので、Entrezによる別名のフラグを結果取得済みに変更する。ついで、出力フラグがすべて結果取得済みかどうかを調べると、BLASTによる名前のフラグが結果取得済みになっていないことがわかる。よって、解析結果142から、名前としてPDS、配列として塩基配列及びアミノ酸配列を抽出し、これら種類とキーワードの組を再検索クエリ151としてクエリ生成部12に送る。

#### 【0108】

クエリ生成部12は、出力制御部15から再検索クエリ151を受け取ると、検索済みフラグと照らし合わせて、検索済みの組み合わせが無いかを調べる。すると、BLAST1フラグのみが未検索であることがわかる。次にクエリ生成部12は、データベース情報を参照し、再検索クエリ151が含む組み合わせで検索可能なデータベースを調べる。すると、BLAST1から、再検索クエリ151が含む配列でBLASTによる相同性検索が可能であることがわかるので、前述の1回目、2回目の検索と同様にして、生物情報データベースクエリ121を生成する。この3回目の生物情報データベースクエリ121は、再検索クエリ151が含む塩基配列あるいはアミノ酸配列でBLASTによる検索を行うものとなる。

#### 【0109】

生物情報データベース検索部13は、1回目、2回目の検索と同様にして検索を行って、生物情報データベース検索結果133として、BLASTによる相同性検索の結果を検索結果解析部14に送る。

## 【0110】

通常のWebブラウザを用いてBLASTによる相同性検索を行うと、図15に示すような画面を得ることができる。図15に示すように、その結果には、入力とした塩基あるいはアミノ酸の配列と相同性のある塩基配列あるいはアミノ酸配列が列挙され、名前、アクセッション番号、相同性の高さを示すe-value、相同性の高い部位等の情報が含まれている。

## 【0111】

BLASTによる配列相同性検索結果について、本実施例の抽出条件161は、e-valueを参照して、相同性の高い配列の遺伝子あるいはタンパク質の名前のみを抽出するようにしておく。これにより、検索結果解析部14は、BLASTによる検索結果から遺伝子あるいはタンパク質の名前を抽出し、解析結果142を出力する。

## 【0112】

出力制御部15は、解析結果142を受け取ってこれを保持するとともに、出力フラグの変更を行う。ここで得た解析結果142は、BLASTによる検索から得た名前を含むので、BLASTによる名前のフラグを取得済みに変更する。ついで、出力フラグがすべて結果取得済みかどうかを調べると、すべての出力フラグが取得済みになっていることがわかるので、出力制御部15は、これまで保持しておいたすべての解析結果142を、出力152として出力する。

## 【0113】

以上では、出力制御部15は、解析結果142を受け取るたびに保持しておき、出力フラグがすべて結果取得済みになってからまとめて出力する例を説明したが、解析結果142を受け取るたびに出力を行ってもよい。また、出力フラグがすべて結果取得済みになったときに、検索完了を示す情報を出力152に加えてもよい。

## 【0114】

図16を参照すると、本発明の第二の実施形態が示されている。本発明の第二の実施形態の生物情報検索システム及び検索方法は、図16のブロック構成図に示すように、第一の実施形態に加えて、出力情報設定部21を有する構成となっ

ている。よって以下では出力情報設定部 21 に関わる説明のみ行う。

#### 【0115】

出力情報設定部 21 は、出力情報指定保持部 16 の保持する検索結果の抽出条件や解析結果の出力条件をユーザによって指定できるようにするものである。これによって、どの生物情報データベースからどの種類のデータを出力させるかをユーザが指定することが可能になる。

#### 【0116】

出力情報設定部 21 は、ユーザ出力情報指定 201 を受け取る。ユーザ出力情報指定 201 は、生物情報データベースと得られる情報の種類を含む。出力情報設定部 21 は、例えば出力情報設定部 21 が情報の種類に応じた条件指定欄を備えることにより、どの生物情報データベースからどのような種類の情報を抽出あるいは出力させるかを決定し、生物情報データベースと抽出あるいは出力させる情報の種類を組にする。このような組が複数あればリスト化し、出力情報設定部 21 は、出力指定情報 211 として出力情報指定保持部 16 に送る。

#### 【0117】

出力情報設定部 21 及びユーザ入力受付部 11 は、どちらもユーザインタフェースとなる部分であり、図 17 のブロック構成図に示すように、これらをまとめた第二のユーザ入力受付部 22 としてもよい。

#### 【0118】

次に、第二の実施形態の具体的な実施例について、図面を参照して詳細に説明する。以下は、ユーザが出力条件として、「Entrezデータベースから遺伝子あるいはタンパク質の別名を取得」、「LocusLinkデータベースから遺伝子あるいはタンパク質の別名を取得」、「BLASTの結果から相同性のある遺伝子あるいはタンパク質の別名を取得」の 3 つを指定する場合を示している。ここで、本発明の第二の実施形態は出力情報設定部 21 を持つ点が、第一の実施形態と異なっている。よって以下では、主に出力情報設定部 21 に関する説明のみ行う。

#### 【0119】

図 18 は、ユーザによるユーザ出力情報指定 201 の入力画面を示している。ユーザが、「Entrez」の「name」、「LocusLink」の「name」、「BLAST」の「na

me」のそれぞれに対応するチェックボックスに印をつけ、submit ボタンをクリックすると、出力情報設定部 21 は、ユーザ出力情報指定 201 として、ユーザが出力条件を「Entrez-遺伝子あるいはタンパク質の別名」、「LocusLink-遺伝子あるいはタンパク質の別名」、「BLAST-相同性のある遺伝子あるいはタンパク質の名前」と指定したと対応付ける。そして、出力情報設定部 21 は、これらの組の情報を出力指定情報 211 として出力情報指定保持部 16 に送る。

#### 【0120】

また、このとき、ユーザが e-value の欄に数値を入力し、submit ボタンをクリックすることによって、BLAST で、相同性検索を行った際に、その結果から、e-value によって、ある程度の相同性を持つ遺伝子あるいはタンパク質のみから、名前を取得することが可能になる。

#### 【0121】

図 19 のブロック構成図を参照すると、本発明の第三の実施形態が示されている。本発明の第三の実施形態の生物情報検索システム及び検索方法は、生物情報データベースの検索手順をあらかじめ決めてから実際の検索を行うものである。

#### 【0122】

図 19 に示すように、本発明の第三の実施形態は、第二のユーザ入力受付部 22 と、検索手順設定部 32 と、生物情報データベース検索制御部 33 と、生物情報データベース検索部 13 と、検索結果解析部 14 と、出力情報指定保持部 16 とを有する。

#### 【0123】

ここで、本発明の第三の実施形態における第二のユーザ入力受付部 22 と、生物情報データベース検索部 13 と、検索結果解析部 14 と、出力情報指定保持部 16 は、第一、第二の実施形態と同じものである。よって以下では、主に検索手順設定部 32 と生物情報データベース検索制御部 33 の動作について説明する。

#### 【0124】

検索手順設定部 32 は、第二のユーザ入力受付部 22 から、ユーザクエリ 111 を受け取る。また、ユーザ出力指定情報 211 を受け取る。

#### 【0125】



検索手順設定部 32 は、各種の生物情報データベース 999 について、その所在、問い合わせ可能な情報（以下「入力」と記す）の種類、及び、検索によって得られる情報（以下「出力」と記す）の種類を組にした情報のリストをデータベース情報として保持している。

#### 【0126】

検索手順設定部 32 は、ユーザクエリ 111 を受け取ると、ユーザクエリ 111 が含む検索キーワードの種類と、データベース情報の各リストが含む入力の種類を比較し、ユーザクエリ 111 の検索キーワードで問合せを行うことができるデータベースを決定する。そして、はじめに問合せを行うデータベースに関して、所在、入力の種類及び出力の種類を組にした情報を生物情報データベースクエリリスト 321 として保持する。

#### 【0127】

次に、検索手順設定部 32 は、この生物情報データベースクエリリスト 321 が含む出力の種類が、出力指定情報 211 で指定する出力の種類に一致するかを調べ、一致する場合は、生物情報データベースクエリリスト 321 とユーザクエリ情報 322 を生物情報データベース検索制御部 33 に送る。一致しない場合は、データベース情報に基づいて、生物情報データベースクエリリスト 321 の最後に保持されている組が含む出力の種類で問合せを行えるデータベースを求め、所在、入力の種類及び出力の種類を組にした情報を生物情報データベースクエリリスト 321 に追加する。このように、生物情報データベースクエリリスト 321 の最後に保持されている、所在、入力の種類及び出力の種類の組について、出力の種類が出力指定情報 211 で指定する出力の種類に一致するまで、上述のような生物情報データベースクエリリスト 321 への追加を繰り返す。出力指定情報 211 で指定する出力の種類に一致する組が見つかったら、ユーザクエリ 111 からそこまでに至る組のリストを改めて生物情報データベースクエリリスト 321 とし、検索手順設定部 32 は、生物情報データベースクエリリスト 321 とユーザクエリ情報 322 を生物情報データベース検索制御部 33 に送る。ユーザクエリ情報 322 は、ユーザ受付部 22 から受け取ったユーザクエリ 111 と同じものである。

**【0128】**

上記では、ユーザクエリ 111 から順次、問い合わせ可能なデータベースを探索し、生物情報データベースクエリリスト 321 を生成する処理を説明したが、出力指定情報 211 で指定するデータベースと出力の種類の情報から、問い合わせ可能なデータベースを順次探索して、ユーザクエリ 111 が含む検索キーワードの種類に一致させることによって生物情報データクエリリスト 321 を生成してもよい。この方法は、問い合わせ可能なデータベースを探索する処理を逆向きに行えば、上記の方法と同様の論理で実現でき、もちろん、上記の方法と組み合わせて、両側から探索してもよい。

**【0129】**

生物情報データベース検索制御部 33 は、検索手順設定部 32 から、生物情報データベースクエリリスト 321 を受け取る。また、ユーザクエリ 322 を受け取る。

**【0130】**

生物情報データベース検索制御部 33 は、まず、生物情報データベースクエリリスト 321 の先頭にある、データベースの所在と入力となるデータの種類、出力となるデータの種類の組を取り出し、ユーザクエリ情報 322 に基づいて、実際に検索を行うための生物情報データベースクエリ 331 を生成する。この生物情報データベースクエリ 331 は、生物情報データベースの URL と検索キーワードを含む、実際の検索要求そのものの形で生成される。生物情報データベース検索制御部 33 は、このようにして生成した生物情報データベースクエリ 331 を生物情報データベース検索部 13 に送る。

**【0131】**

生物情報データベース検索部 13 及び検索結果解析部 14 における検索処理及び検索結果の解析処理は、第一及び第二の実施形態におけるこれらの構成要素が行う処理と同様である。また、出力情報指定保持部 16 の処理も第一及び第二の実施形態のそれと同様である。

**【0132】**

生物情報データベースの検索と検索結果の解析の後、生物情報データベース検

索制御部 33 は、検索結果解析部 14 から解析結果 142 を受け取り、生物情報データベースクエリリスト 321 の先頭にある組の内容に基づいて、解析結果 142 から、検索を行ったデータベースの出力の種類及び内容を抽出して保持する。

#### 【0133】

このようにして 1 回目のデータベース検索を行った後、生物情報データベース検索制御部 33 は、生物情報データベースクエリリスト 321 が含む手順の検索をすべて行ったかどうかを判定する。すべての検索が済んだ場合は、保持している出力情報の種類と内容の情報を出力 152 として出力する。まだ生物情報データベースクエリリスト 321 が含む手順の検索が済んでいない場合は、生物情報データベースクエリリスト 321 から次に検索すべき手順の情報を取り出し、直前の手順による検索の出力から得た情報から新たな検索キーワードを抽出して、新たな生物情報データベースクエリ 331 を生成する。生物情報データベース検索制御部 33 は、新しいデータベースクエリ 331 を生物情報データベース検索部 34 に送り、以降、生物情報データベースクエリリスト 321 の最後の組に基づく検索が完了するまで、同様の処理を繰り返す。

#### 【0134】

図 20 のフローチャートを参照すると、検索手順設定部 32 の処理の詳細が示されている。

#### 【0135】

検索手順設定部 32 は、ステップ 3201 で、第二のユーザ入力受付部 22 からユーザクエリ 111 を読み込む。また、ステップ 3202 では、第二のユーザ入力受付部 22 から出力情報指定 211 を読み込む。続いて、ステップ 3203 では、データベース情報として保持している、データベースの種類、及び入力の種類、出力の種類の情報を読み込む。

#### 【0136】

次に、ステップ 3204 では、ユーザクエリ 111 が含むデータの種類の、データベース情報が含むデータの输入の種類を比較し、ユーザクエリ 111 の含む検索キーワードで問い合わせのできるデータベースの所在を決定する。そして、

問い合わせを行うデータベースに関して、所在、入力の種類及び出力の種類の情報組にして、生物情報データベースクエリリスト 321 として保持する。

#### 【0137】

そして、ステップ 3205 では、生物情報データベースクエリリスト 321 が含むデータベースの種類と出力の種類の情報と、出力指定情報 211 が含むデータベース及びデータの種類組と比較する。もし、生物情報データベースクエリリスト 321 が出力指定情報 211 を満たした場合、ステップ 3206 において、保持していた生物情報データベースクエリリスト 321 を生物情報データベース検索制御部 33 に送る。また、ステップ 3207 で、ユーザクエリ 111 の検索キーワードとその種類組を、そのままユーザクエリ情報 322 として生物情報データベース検索部 33 に送る。もし、生物情報データベースクエリリスト 321 が出力指定情報 211 を満たさない場合、ステップ 3208 において、生物情報データベースクエリリスト 321 の最後に保持されている組が含む、出力の種類情報を抽出し、ステップ 3209 では、それを保持する。

#### 【0138】

ここで、ステップ 3204 において、複数の生物情報データベースクエリリスト 321 が生成されるが、そのうちひとつでもステップ 3205 の条件を満たせば、残りの生物情報データベースクエリリスト 321 については、処理を行わない。もし、すべての生物情報データベースクエリリスト 321 が、ステップ 3205 の条件を満たさなかったとき、保持していた出力の種類情報をステップ 3204 に戻し、問い合わせを行えるデータベースを求め、所在、入力の種類、出力の種類組にして、生物情報データベースクエリリスト 321 に追加する。このように、生物情報データベースクエリリスト 321 の最後に保持されている、所在、入力の種類、出力の種類組について、出力の種類が出力指定情報 211 で指定するデータベースと出力の種類に一致するまで、上述のような生物情報データベースクエリリスト 321 への追加を繰り返す。もちろん、生物情報データベースクエリリスト 321 への追加の回数をどの段階まで行うかを規定してもよく、また、ユーザが指定できるようにしてもよい。

#### 【0139】

上記では、ユーザクエリ 111 から順次、問い合わせ可能なデータベースを探索し、生物情報データベースクエリリスト 321 を生成する処理を説明したが、出力指定情報 211 で指定するデータベースと出力の種類の情報から、問い合わせ可能なデータベースを順次探索して、ユーザクエリ 111 が含む検索キーワードの種類に一致させることによってクエリリスト 321 を生成してもよい。この方法は、問い合わせ可能なデータベースを探索する処理を逆向きに行えば、上記の方法と同様の論理で実現でき、もちろん、上記の方法と組み合わせて、両側から探索してもよい。

#### 【0140】

図 21 のフローチャートを参照すると、生物情報データベース検索制御部 33 の処理の詳細が示されている。

#### 【0141】

ステップ 3301 では、生物情報データベースクエリリスト 321 を検索手順設定部 32 から受け取る。続いて、ステップ 3302 では、検索手順設定部 32 からユーザクエリ情報 322 を受け取る。

#### 【0142】

ステップ 3303 では、生物情報データベースクエリリスト 321 の先頭にあるデータベースの所在、入力の種類、出力の種類の組を取り出し、保持する。

#### 【0143】

次に、ステップ 3304 では、取り出した組と、ユーザクエリ情報 322 の検索キーワードとその種類の情報から、生物情報データベースクエリ 331 を生成する。

#### 【0144】

そして、ステップ 3305 では、生成した生物情報データベースクエリ 331 を生物情報データベース検索部 13 に送る。

#### 【0145】

生物情報データベース検索部 13 による生物情報データベース 999 の検索と、検索結果解析部 14 での検索結果の解析の後、ステップ 3306 では、検索結果解析部 14 から、解析結果 142 を受け取り、それを出力情報 152 として保

持する。

#### 【0146】

次にステップ3207では、生物情報データベースクエリリスト321が含む手順の検索をすべて行ったか判定する。すべての検索が済んだ場合は、ステップ3208において、保持している出力情報152を出力する。まだ、すべての検索が済んでいない場合は、保持している出力情報152をステップ3303に送る。そして、再び検索すべき手順の情報を取り出し、新たな生物情報データベースクエリ331を生成する。ここで、ステップ3304では、ユーザクエリ322でなく出力情報152に基づいて、生物情報データベースクエリ331を生成する。以降、生物情報データベースクエリリスト321の最後の組に基づく検索が完了するまで、同様の処理を繰り返す。

#### 【0147】

次に、第三の実施形態の具体的な実施例について、図面を参照して詳細に説明する。以下においても、第一の実施の形態の具体的な実施例と同様、ユーザが遺伝子あるいはタンパク質の名前の検索キーワードとして「pendrin」と入力し、出力条件として「Entrez」の「name」、「LocusLink」の「name」、「BLAST」の「name」のそれぞれに対応するチェックボックスに印をつけたときに、EntrezとLocusLinkからはその別名を取得し、また、Entrezで別名検索を行う過程で得た配列情報を用いてBLASTで相同性検索を行って、その結果から相同性の高い遺伝子あるいはタンパク質の名前を取得する場合を示している。

#### 【0148】

ここで、本発明の第三の実施形態における第二のユーザ入力受付部22と、生物情報データベース検索部13と、検索結果解析部14と、出力情報指定保持部16は、第一、第二の実施形態と同じものである。よって以下では、主に検索手順設定部32と生物情報データベース検索制御部33の動作について説明する。

#### 【0149】

検索手順設定部32は、ユーザクエリ111として、検索キーワードが「pendrin」であり、その種類が遺伝子あるいはタンパク質の名前であるという情報を受け取る。また、出力指定情報211として、出力条件が「Entrez-遺伝子ある

いはタンパク質の別名」、「LocusLink-遺伝子あるいはタンパク質の別名」、「BLAST-相同性のある遺伝子あるいはタンパク質の名前」であるという情報を受け取る。

### 【0150】

続いて、データベース情報として保持している、データベースの種類、及び入力の種類、出力の種類の情報を読み込む。本実施例では、データベース情報は、{データベースの種類、受け付ける入力の種類、得られる出力の種類}として、次のような情報を持つものとする。

Entrez 1 1 = {Entrezの所在 1、遺伝子あるいはタンパク質の名前、遺伝子あるいはタンパク質の別名}

Entrez 1 2 = {Entrezの所在 2、遺伝子あるいはタンパク質の名前、GI番号}

LocusLink 1 1 = {LocusLinkの所在 1、遺伝子あるいはタンパク質の名前、遺伝子あるいはタンパク質の別名}

BLAST 1 1 = {BLASTの所在 1、配列情報、相同性のある遺伝子あるいはタンパク質の名前}

ここで、「Entrezの所在 1」とは、EntrezそのもののURLにCGIによる検索文字列のテンプレートを結合したものとする。テンプレートとは、検索キーワード部分のみが置換可能な変数名に置き換えられたものである。「Entrezの所在 2」、「LocusLinkの所在 1」、「BLASTの所在 1」等も同様のものとする。もちろん、URLはローカルなデータベースの所在も指定できる。

### 【0151】

ユーザクエリ 1 1 1によると、ユーザの入力によって、遺伝子あるいはタンパク質名が与えられているので、上記のようなデータベース情報から、検索手順設定部 3 2では、Entrez 1 1、Entrez 1 2及び、LocusLink 1 1を用いて検索を行うことを決定する。そして、生物情報データベースクエリリスト 3 2 1として保持する。生物情報データベースクエリリスト 3 2 1は、次のような情報を持つものとする。

問合せ (1、1) = {Entrezの所在 1、遺伝子あるいはタンパク質の名前、遺伝子あるいはタンパク質の別名}

問合せ (1, 2) = {Entrezの所在 2、遺伝子あるいはタンパク質の名前、GI番号}

問合せ (1, 3) = {LocusLinkの所在 1、遺伝子あるいはタンパク質の名前、遺伝子あるいはタンパク質の別名}

そして、生物情報データベースクエリリスト 3 2 1 が含むデータベースの種類と出力の種類の情報と、出力指定情報 2 1 1 が含むデータベース及びデータの種類、この場合、LocusLinkの別名、Entrezの別名は出力指定情報 2 1 1 を満たすが、まだ、BLASTによる名前の条件は満たさない。よって、生物情報データベースクエリリスト 3 2 1 の最後の組が含む出力の種類、この場合、遺伝子あるいはタンパク質の別名とGI番号という情報を抽出する。この出力の種類を新たな入力の情報として、データベース情報と照らし合わせると、新たな問合せの組が決定される。この場合、次のような組等が決定される。

#### 【0152】

問合せ (2, 1) = {Entrezの所在 2、GI番号、塩基あるいはアミノ酸配列}

そして、この組の情報を生物情報データベースクエリリスト 3 2 1 に追加する。このとき、生物情報データベースクエリ 3 2 1 は、次のような状況になっている。

{問合せ (1, 1) }

{問合せ (1, 2)、問合せ (2, 1) }

{問合せ (1, 3) }

ここで再び、生物情報データベースクエリリスト 3 2 1 が含むデータベースの種類と出力の種類の情報と、出力指定情報 2 1 1 が含むデータベース及びデータの種類、この場合、まだBLASTによる名前の条件は満たさない。よって、生物情報データベースクエリリスト 3 2 1 の最後の組が含む出力の種類、この場合、塩基あるいはアミノ酸配列という情報を抽出する。この出力の種類を新たな入力の情報として、データベース情報と照らし合わせると、新たな問合せの組が決定される。この場合、次のような組等が決定される。



問合せ (3, 1) = {BLASTの所在 1、塩基あるいはアミノ酸配列、遺伝子あるいはタンパク質の名前、遺伝子あるいはタンパク質の別名}

そして、この組の情報を生物情報データベースクエリリスト 3 2 1 に追加する。このとき、生物情報データベースクエリ 3 2 1 は、次のような状況になっている。

{問合せ (1, 1) }

{問合せ (1, 2)、問合せ (2, 1)、問合せ (3, 1) }

{問合せ (1, 3) }

ここで再び、生物情報データベースクエリリスト 3 2 1 が含むデータベースの種類と出力の種類の情報と、出力指定情報 2 1 1 が含むデータベース及びデータの種類と組と比較する。すると、この場合は、すべての条件を満たすので、保持していた生物情報データベースクエリ 3 2 1 を生物情報データベース検索制御部 3 3 に送る。また、ユーザクエリ 1 1 1 の検索キーワードが pendrin で、その種類が遺伝子あるいはタンパク質の名前であるという情報をユーザクエリ情報 3 2 2 として生物情報データベース検索制御部 3 3 に送る。

#### 【0153】

上記では、ユーザクエリ 1 1 1 から順次、問合せ可能なデータベースを探索し、生物情報データベースクエリリスト 3 2 1 を生成する処理を説明したが、出力指定情報 2 1 1 で指定するデータベースと出力の種類の情報から、問い合わせ可能なデータベースを順次探索して、ユーザクエリ 1 1 1 が含む検索キーワードの種類に一致させることによってクエリリスト 3 2 1 を生成してもよい。この方法は、問い合わせ可能なデータベースを探索する処理を逆向きに行えば、上記の方法と同様の論理で実現でき、もちろん、上記の方法と組み合わせて、両側から探索してもよい。

#### 【0154】

生物情報データベース検索制御部 3 3 は、生物情報データベースクエリリスト 3 2 1 と、ユーザクエリ情報 3 2 2 として検索キーワードが pendrin で、その種類が遺伝子あるいはタンパク質の名前であるという情報を受け取る。

#### 【0155】

次に、生物情報データベースクエリリスト 321 の先頭にあるデータベースの種類、入力の種類、出力の種類の組を取り出し、保持する。今回の場合、例えば、問合せ (1, 1) = {Entrezの所在 1、遺伝子あるいはタンパク質の名前、遺伝子あるいはタンパク質の別名} である。そして、これと検索キーワードが pendrin で、その種類が遺伝子あるいはタンパク質の名前であるという情報から、実際に検索を行う生物情報データベースクエリ 331 を生成する。ここで生成された、生物情報データベースクエリ 331 は、Entrez そのものの URL に CGI による検索文字列を結合したものであり、検索文字列のテンプレート中の検索キーワードが pendrin に置き換えられたものである。生物情報データベース検索制御部 33 は、生物情報データベースクエリ 331 を生物情報データベース検索部 13 に送る。すると、検索が行われ、その結果が検索結果解析部 14 で解析される。そして、検索結果解析部 14 は、解析結果 142 を生物情報データベース検索制御部 33 に送る。

生物情報データベース検索制御部 33 は、解析結果 142 を受け取ると、それを出力情報 152 として保持する。ここには、Entrez から得た別名として PDS、GI 番号として 265004 等の情報が含まれている。

#### 【0156】

次に、生物情報データベースクエリリスト 321 が含む手順の検索をすべて行ったか判定するが、まだすべては行っていない。そこで生物情報データベースクエリリストから次の組を切り出し、生物情報データベース 999 の検索を行う。

#### 【0157】

LocusLink における別名の取得の処理も同様に行う。すると、別名として SLC26A4、DFNB4、PDS、pendrin 等の情報が得られる。次に、生物情報データベースクエリリスト 321 が含む手順の検索をすべて行ったか判定するが、まだすべては行っていない。

#### 【0158】

前述と同様、Entrez を遺伝子あるいはタンパク質の名前として pendrin で検索し、GI 番号 265004 を得る。次に手順に従い、Entrez を GI 番号として 265004 で検索すると、塩基あるいはアミノ酸配列が得られる。この塩基あるいはアミノ酸配

列をBLASTで相同性検索を行って、その結果より、相同性のある遺伝子あるいはアミノ酸の名前が得られる。ここで、生物情報データベースクエリリスト321が含む手順の検索をすべて行ったかを判定すると、すべて行ったので、保持していた出力152を出力する。

#### 【0159】

図22のブロック構成図を参照すると、本発明の第四の実施形態が示されている。本発明の第四の実施形態の生物情報検索システム及び検索方法は、前述の第一乃至第三の実施形態における出力制御部及び生物情報データベース検索制御部が出力する情報を受け取り、遺伝子名及びタンパク質名の補完を行って出力するものである。

#### 【0160】

図22に示すように、本発明の第四の実施形態は、前述の第一乃至第三の実施形態に検索キーワード補完部41が追加された構成となっている。よって、以下では、検索キーワード補完部41についてのみ説明する。

#### 【0161】

検索キーワード補完部41は、遺伝子あるいはタンパク質の名前に関わる所定の書式と補完ルールを持っている。そして、出力152を受け取ると、出力152が含む遺伝子あるいはタンパク質の名前について、それらが、英字及び数字から成る所定の形式である場合は、補完ルールに従って別の表記のしかたによるものを生成して出力152に追加し、第二の出力411として出力する。

#### 【0162】

次に、第四の実施形態の具体的な実施例について、詳細に説明する。

#### 【0163】

例えば、検索キーワード補完部41が出力152を受け取ったとき、出力152が遺伝子あるいはタンパク質の名前として、「CA1」を含んでいた場合の例を示す。ここでは、遺伝子あるいはタンパク質の名前に関わる所定の書式として、英字が複数文字続いた後に数字という書式を保持しているものとする。また、補完ルールとして、英字と数字の間にスペース、ハイフンを補い、アラビア数字はローマ数字で表記したものを補完するというルールを設定しておくものとする。

。これらの書式及び補完ルールにより、出力 152 に含まれる「CA 1」について、検索キーワード補完部 41 は、「CA 1」「CA-1」「CA I」「CA-I」等を出力 152 に付け加える。そして、検索キーワード補完部 41 は、これを第二の出力 411 として出力する。

#### 【0164】

また、遺伝子あるいはタンパク質の名前に関する書式として、英字が複数続いた後にスペースをはさんで数字という書式を保持し、補完ルールとして、英字と数字の間のスペースを削除する、スペースをハイフンに置換する、数字部分はアラビア数字とローマ数字の両方となるよう補完する等のルールを設定しておけば、出力 152 が遺伝子あるいはタンパク質の名前として「CA 1」を含んでいる場合、検索キーワード補完部 41 は、「CA 1」「CA-1」「CA I」「CA-I」等を出力 152 に付け加える。そして、検索キーワード補完部 41 は、これを第二の出力 411 として出力する。

#### 【0165】

なお、以上のように遺伝子あるいはタンパク質の名前の補完を行えば、出力 152 が遺伝子あるいはタンパク質の名前として、「solute carrier family 26, member 4」を含んでいた場合や、「p53」を含んでいた場合には、保持している書式に該当しないので、検索キーワード補完部 41 は、特に補完を行わずに、出力 152 をそのまま第二の出力 411 として出力する。つまり、上述のような書式及び補完ルールを設定することにより、「solute carrier family 26, member 4」や「p53」のように、補完すべきでない場合にはそのまま出力させることができる。

#### 【0166】

図 23 のブロック構成図を参照すると、本発明の第五の実施形態が示されている。本発明の第五の実施形態の生物情報検索システム及び検索方法は、前述の第一乃至第四の実施形態における出力制御部、生物情報データベース検索制御部及び検索キーワード補完部が出力する出力 152 あるいは第二の出力 411 を受け取り、それを用いて学术论文あるいは特許等の文献検索を行うものである。

#### 【0167】

図 23 に示すように、本発明の第五の実施形態は、前述の第一乃至第四の実施形態に文献データベース検索部 51 が追加された構成となっている。よって、以下では、文献データベース検索部 51 についてのみ説明する。

#### 【0168】

文献データベース検索部 51 は、各種の文献データベース 1000 について、その所在の情報を保持している。ここで、データベースの所在は、例えば URL で表すことができる。また、文献データベース検索部 51 は、文献データベース 1000 に示される各種の文献データベースから得られる検索結果に応じた文献データベース検索結果抽出条件を保持している。文献データベース検索結果抽出条件は、各文献データベースの検索結果について、出力に用いる部分を特定するための情報や抽出のルール、文献の一覧表示から各文献の詳細情報を抽出するルール、検索結果が得られない場合に各文献データベースが返す情報の書式等を含んでいる。

#### 【0169】

文献データベース検索部 51 は、出力 152 あるいは第二の出力 411 を受け取ると、出力 152 あるいは第二の出力 411 が含む遺伝子名、タンパク質名、及びこれらの別名等を文献検索キーワードとし、それらの種類とともに組にして保持する。各組について、文献データベース検索部 51 は、保持している文献データベースの所在の情報に基づき、文献データベース 1000 を検索するための文献データベース問合せ 511 を生成し、順次、文献データベース 1000 に送る。文献データベース 1000 から文献データベース検索結果 512 を受け取るごとに、文献データベース検索部 51 は、文献データベース検索結果抽出条件に基づき、文献データベース検索結果 512 から、雑誌名、著者名あるいは発明者名、発行日、本文、各文献データベースがつけた ID 番号等のデータを抽出して文献検索結果 513 として保持する。さらに、文献データベース検索部 51 は、文献データベース検索結果抽出条件に基づいて、文献データベース検索結果 512 から各文献の詳細情報を検索する情報が抽出可能かどうかを判定し、可能な場合は、詳細情報を得るための文献データベース検索を行って、その結果を文献検索結果 513 に追加して保持する。これらの処理を、出力 152 あるいは第二の出

力 4 1 1 に含まれるすべての文献検索キーワードで文献データベース 1 0 0 0 を検索し終わるまで繰り返す。出力 1 5 2 あるいは第二の出力 4 1 1 に含まれるすべての文献検索キーワードで文献データベース 1 0 0 0 を検索し終わった場合、文献データベース検索部 5 1 は保持していた文献検索結果 5 1 3 を出力する。

#### 【0170】

図 2 4 のフローチャートを参照すると、文献データベース検索部 5 1 の処理の詳細が示されている。

#### 【0171】

文献データベース検索部 5 1 は、出力 1 5 2 あるいは第二の出力 4 1 1 を受け取る。

#### 【0172】

ステップ 5 1 0 1 では、保持している文献データベースの所在の情報を読み込む。続いて、ステップ 5 1 0 2 では、保持している文献データベース検索結果抽出条件を読み込む。次にステップ 5 1 0 3 では、受け取った出力 1 5 2 あるいは第二の出力 4 1 1 の先頭から、文献検索キーワードとその種類の組を抽出する。そして、ステップ 5 1 0 4 では、抽出した組の情報と、保持していた文献データベースの所在の情報をもとに、文献検索問合せ 5 1 1 を生成し、ステップ 5 1 0 5 で、文献検索問合せ 5 1 1 を文献データベース 1 0 0 0 に送る。ステップ 5 1 0 6 では、文献データベース 1 0 0 0 から、文献データベース検索結果 5 1 2 を受け取り、それを保持する。次に、ステップ 5 1 0 7 では、保持している文献データベース検索結果抽出条件に基づき、文献データベース検索結果 5 1 2 から、雑誌名、著者名あるいは発明者名、発行日、本文、各文献データベースがつけた ID 番号等のデータを抽出し、それを文献検索結果 5 1 3 として保持する。

#### 【0173】

ステップ 5 1 0 8 では、文献データベース検索結果 5 1 2 から各文献の詳細情報を検索する情報が抽出可能かどうかを判定し、可能な場合は、詳細情報を得るための文献検索問合せ 5 1 1 を生成し、ステップ 5 1 0 3 からの検索処理を再び行わせる。文献の詳細情報を検索する情報が抽出できない場合は、ステップ 5 1 0 9 へと進む。ステップ 5 1 0 9 では、出力 1 5 2 あるいは第二の出力 4 1 1 と

して受け取った文献検索キーワードによる検索をすべて行ったか確かめる。すべて行った場合、ステップ5110で、保持していた文献検索結果513を出力する。まだすべて検索が終わっていない場合は、ステップ5103に戻る。

#### 【0174】

ここで、ステップ5109で、文献検索結果513を保持する代わりに逐次出力をしてもよい。その場合、ステップ5110では、文献検索を終了する。また、ここで文献検索を終了する代わりに、検索完了を示す情報を出力してもよい。これにより、利用者に検索完了を明示することができる。

#### 【0175】

また、上記では、文献データベース1000の所在や、文献データベース検索結果512からどの文献データを抽出するかは規定であったが、もちろん、これらをユーザが指定することもできる。

#### 【0176】

次に、第五の実施形態の具体的な実施例について、図面を参照して詳細に説明する。以下は、ユーザが遺伝子あるいはタンパク質の名前の検索キーワードとして「pendrin」と入力し、PubMedにより学術論文の文献検索を行った場合を示している。

#### 【0177】

ユーザが「pendrin」を入力としたとき、前述の第一乃至第四の実施形態により、別名として、「PDS」、「DFNB4」、「SLC26A6」等が得られている。文献データベース検索部51はこの情報を受け取る。文献データベース検索部51は、各種の文献データベース1000について、その所在の情報を保持している。そして、文献データベース検索部51は、データベースの所在の情報と、文献検索キーワードがpendrinという情報から、文献検索問合せ511を生成する。ここで生成された文献検索問合せ511は、本実施例では、PubMedそのもののURLにCGIによる検索文字列を結合したものであり、検索文字列のテンプレート中の検索キーワードがpendrinに置き換えられたものである。

#### 【0178】

通常のWebブラウザを用いた検索において、PubMedで遺伝子あるいはタンパク

質の名前としてpendrinを検索すると、図25に示すような結果が得られる。図25に示すように、その結果は、所定の場所に「Hulander M」といった著者名、「Lack of pendrin expression leads to deafness and expansion of the endolymphatic compartment in inner ears of Foxil null mutant mice」といった題名、「Development」といった雑誌名、「2003 May 1」といった発行日、「12642503」といったPubMed ID等が表示される。

#### 【0179】

これらのことから、文献データベース検索の結果として、通常のWebブラウザによる検索と同様の結果が得られるならば、HTML形式の検索結果中、所定位置もしくは書式の部分を抽出すれば、著者名や題名あるいは雑誌名等の情報が得られる。この抽出した情報を文献検索結果513として保持する。

#### 【0180】

次に、この場合、文献データベース検索部51は、図25に示すような文献データベース検索結果512であるので、その書式から各文献の詳細情報が検索可能であり、そのための情報が抽出可能であると判定する。そこで、文献データベース検索部51は、PubMed IDの情報を抽出し、各文献の詳細情報を取得するべく、再び、文献データベースの検索を行う。

#### 【0181】

PubMed IDの情報をもとに、前述した処理と同様の処理を行うと、図26に示すような文献検索結果513が得られる。図26に示すように、その結果は、所定の場所に文献の概要、著者名、題名、雑誌名等が表示される。これらのことから、文献データベース検索の結果として、通常のWebブラウザによる検索と同様の結果が得られるならば、文献の概要等の情報が得られる。この抽出した情報を文献検索結果513として保持する。

#### 【0182】

ここで、文献データベース検索部51は、図26に示すような文献データベース検索結果512であるので、その書式から、これ以上の詳細情報は検索できないと判定する。

#### 【0183】



次に、出力 152 あるいは第二の出力 411 として受け取った文献検索キーワードによる検索をすべて行ったか確かめる。今回の場合、検索を行ったのは pen drin に関してのみなので、順次、前述の処理を行って、「PDS」、「DFNB4」、「SLC26A6」等に関する論文を取得する。

#### 【0184】

そして、出力 152 あるいは第二の出力 411 として受け取った文献検索キーワードによる検索をすべて行った後、保持していた文献検索結果 513 を出力する。

#### 【0185】

図 27 のブロック構成図を参照すると、本発明の第六の実施形態が示されている。本発明の第六の実施形態の生物情報検索システム及び検索方法は、第五の実施形態で得た文献検索結果を統合したり、並べ替えたりする等により出力するものである。

#### 【0186】

図 27 に示すように、本発明の第六の実施形態は、上述の第五の実施形態に、文献検索結果解析部 61 が加わった構成である。よって、以下では、文献検索結果解析部 61 についてのみ説明する。

#### 【0187】

文献検索結果解析部 61 は、文献検索結果解析ルールを保持している。文献検索結果解析ルールは、例えば、文献検索結果 513 が同一の文献に関する同一の情報を含む場合に重複を除去する、文献ごとにインデックスを付ける、文献数を数える、発表年月の順に並べ替える、雑誌名で並べ替える、著者名で並べ替える等のルールを含む。

#### 【0188】

文献検索結果解析部 61 は、文献検索結果 513 を受け取ると、上述のような文献検索結果解析ルールに従って、内容の並べ替え、インデックスの付加、文献数の付加、重複の除去等を行って、文献検索解析結果 611 として出力する。

#### 【0189】

次に、第六の実施形態の具体的な実施例について、図面を参照して詳細に説明

する。以下は、ユーザが遺伝子あるいはタンパク質の名前の検索キーワードとして「pendrin」と入力し、第四の実施形態により、遺伝子あるいはタンパク質の名前の補完を行って、第五の実施形態によりPubMedにより学术论文の文献検索を行った後、「pendrin」と「SLC26A4」に関する学术论文について、重複するものを除去し、結果を統合して表示する場合の例を示す。

#### 【0190】

文献検索結果解析部61は、文献検索結果513を受け取る。図28は、文献検索結果513として、PubMedで「pendrin」を検索した際の結果を示している。また、図29は、文献検索結果として、PubMedにおいて、第四の実施形態により補完された遺伝子あるいはタンパク質の名前である「SLC26A4」を検索した際の結果を示している。すると、各々の文献検索結果513中に、重複する学术论文があることがわかる。そこで、文献検索結果解析部61は、図30のように、重複する論文を除去し、結果を統合した文献検索結果を生成する。

#### 【0191】

図31のブロック構成図を参照すると、本発明の第七の実施形態が示されている。本発明の第七の実施形態の生物情報検索システム及び検索方法は、第五の実施形態あるいは第六の実施形態の出力を受けて、文献検索の結果を整形し出力するものである。

#### 【0192】

図31に示すように、本発明の第七の実施形態は、第五の実施形態あるいは第六の実施形態に整形出力部71が付け加えられた構成である。よって、以下では、整形出力部71についてのみ説明する。

#### 【0193】

整形出力部71は、文献検索結果が含む情報をどのような形式で出力するかを表す整形出力ルールを保持している。整形出力ルールとしては、例えば、はじめにユーザが指定した検索キーワードの表示色を変える、遺伝子名とその別名をまとめて表示する、タンパク質名とその別名をまとめて表示する、すべての文献のID番号と概要をリスト形式で表示する等を定めておく。

#### 【0194】

整形出力部 7 1 は、文献検索結果 5 1 3 あるいは文献検索解析結果 6 1 1 を受け取ると、上述のような整形出力ルールに従って、文献中の検索キーワードの表示色を変える、遺伝子あるいはタンパク質の名前と別名の表示、文献の ID 番号と概要のリスト状の表示等といった出力の整形を行って、整形出力結果 7 1 1 とし

#### 【0 1 9 5】

次に、第七の実施形態の具体的な実施例について、図面を参照して詳細に説明する。以下は、ユーザが遺伝子あるいはタンパク質の名前の検索キーワードとして「pendrin」と入力し、第四の実施形態により、遺伝子あるいはタンパク質の名前の補完を行って、第五の実施形態により PubMed により 学術論文の文献検索を行った後、文献中の検索キーワードの表示色の変更、文献の概要のリスト状の表示等を行った場合の例を示す。

#### 【0 1 9 6】

整形出力部 7 1 は、文献検索結果 5 1 3 あるいは文献検索解析結果 6 1 1 を受け取ると、上述のような整形出力ルールに従って、文献中の検索キーワードの表示色を変える、遺伝子あるいはタンパク質の名前と別名の表示、文献の ID 番号と概要のリスト状の表示等といった出力の整形を行って、図 3 2 のような整形出力結果 7 1 1 を出力する。

#### 【0 1 9 7】

##### 【発明の効果】

以上、説明したように、本発明においては、次のような効果を奏する。

#### 【0 1 9 8】

第一の効果は、検索を行う研究者の知る遺伝子あるいはタンパク質の名前だけで検索を行っても、漏れの少ない検索結果が得られることである。その理由は、本発明の生物情報検索システム及び検索方法が、研究者の知る情報で検索を行った結果を用いて更なる検索を行うからである。

#### 【0 1 9 9】

第二の効果は、遺伝子あるいはタンパク質の名前以外を検索キーワードにして検索を行っても、これらの名前以外で検索できない生物情報データベースからの

検索まで得られることである。その理由は、本発明の生物情報検索システム及び検索方法が、遺伝子あるいはタンパク質の名前以外を検索キーワードにして検索を行って、得られた結果から遺伝子あるいはタンパク質の名前を抽出してさらなる検索を行うからである。

### 【0200】

第三の効果は、研究者が生物情報データベースを研究する際の負担を軽減することである。その理由は、第一及び第二の効果により、研究者自身が生物情報データベースの検索を繰り返し行うことが不要になるからである。

#### 【図面の簡単な説明】

##### 【図1】

本発明の第一の実施形態の生物情報検索システム及び検索方法の構成を示したブロック構成図である。

##### 【図2】

ユーザ入力受付部において実施される処理の流れを説明したフローチャートである。

##### 【図3】

クエリ生成部において実施される処理の流れを説明したフローチャートである。

##### 【図4】

生物情報データベース検索部において実施される処理の流れを説明したフローチャートである。

##### 【図5】

生物情報データベース検索部の別の形態において実施される処理の流れを説明したフローチャートである。

##### 【図6】

検索結果解析部において実施される処理の流れを説明したフローチャートである。

##### 【図7】

出力制御部において実施される処理の流れを説明したフローチャートである。

**【図 8】**

出力制御部の別の形態において実施される処理の流れを説明したフローチャートである。

**【図 9】**

出力制御部のさらなる別の形態において実施される処理の流れを説明したフローチャートである。

**【図 10】**

出力情報指定保持部において実施される処理の流れを説明したフローチャートである。

**【図 11】**

生物情報検索における生物情報の入力画面の一例である。

**【図 12】**

生物情報検索におけるLocusLinkの結果画面の一例である。

**【図 13】**

生物情報検索におけるEntrezの結果画面の一例である。

**【図 14】**

生物情報検索におけるEntrezの結果画面の別の一例である。

**【図 15】**

生物情報検索におけるBLASTの相同性検索結果画面の一例である。

**【図 16】**

本発明の第二の実施形態の生物情報検索システム及び検索方法の構成を示したブロック構成図である。

**【図 17】**

本発明の第二の実施形態の別の形態における生物情報検索システム及び検索方法の構成を示したブロック構成図である。

**【図 18】**

生物情報検索における出力条件の入力画面の一例である。

**【図 19】**

本発明の第三の実施形態の生物情報検索システム及び検索方法の構成を示した

ブロック構成図である。

【図 20】

検索手順設定部において実施される処理の流れを説明したフローチャートである。

【図 21】

生物情報データベース検索制御部において実施される処理の流れを説明したフローチャートである。

【図 22】

本発明の第四の実施形態の生物情報検索システム及び検索方法の構成を示したブロック構成図である。

【図 23】

本発明の第五の実施形態の生物情報検索システム及び検索方法の構成を示したブロック構成図である。

【図 24】

文献データベース検索部において実施される処理の流れを説明したフローチャートである。

【図 25】

文献検索におけるPubMedの結果画面の一例である。

【図 26】

文献検索におけるPubMedの結果画面の別の一例である。

【図 27】

本発明の第六の実施形態の生物情報検索システム及び検索方法の構成を示したブロック構成図である。

【図 28】

遺伝子あるいはタンパク質の名前としてpendrinを入力した際の文献検索結果画面である。

【図 29】

遺伝子あるいはタンパク質の名前としてSLC26A4を入力した際の文献検索結果画面である。

**【図 30】**

遺伝子あるいはタンパク質の名前としてpendrin、SLC26A4を入力した際に、その結果を冗長性をなくす等、検索結果を統合した解析結果画面である。

**【図 31】**

本発明の第七の実施形態の生物情報検索システム及び検索方法の構成を示したブロック構成図である。

**【図 32】**

遺伝子あるいはタンパク質の名前で文献を検索した際に、その結果をルールに基づいて整形した出力結果画面である。

**【符号の説明】**

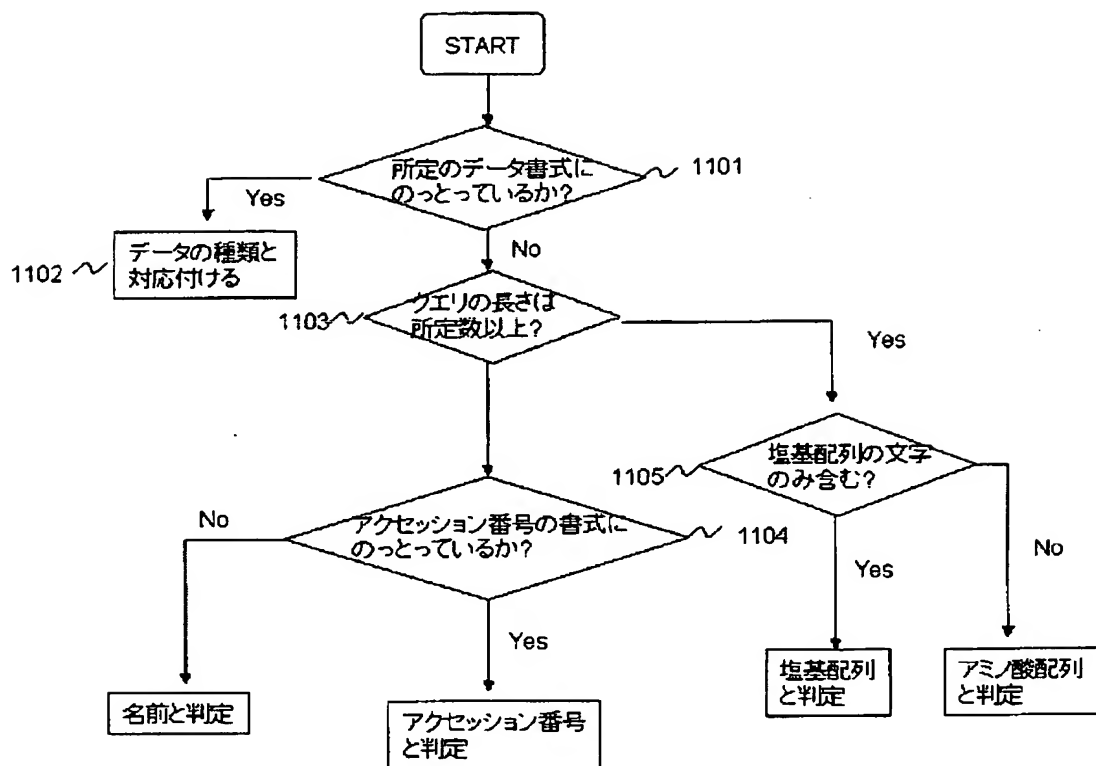
- 1 1 ユーザ入力受付部
- 1 2 クエリ生成部
- 1 3 生物情報データベース検索部
- 1 4 検索結果解析部
- 1 5 出力制御部
- 1 6 出力情報指定保持部
- 2 1 出力情報設定部
- 2 2 第二のユーザ入力受付部
- 3 2 検索手順設定部
- 3 3 生物情報データベース検索制御部
- 4 1 検索キーワード補完部
- 5 1 文献データベース検索部
- 6 1 文献検索結果解析部
- 7 1 整形出力部
- 1 0 1 ユーザ入力
- 1 1 1 ユーザクエリ
- 1 2 1 生物情報データクエリ
- 1 2 2 検索完了情報
- 1 3 1 問合せ

- 1 3 2 検索結果
- 1 3 3 生物情報データベース検索結果
- 1 3 4 エラー情報
- 1 4 1 抽出条件要求
- 1 4 2 解析結果
- 1 5 1 再検索クエリ
- 1 5 2 出力
- 1 6 1 抽出条件
- 1 6 2 出力情報指定
- 2 0 1 ユーザ出力情報指定
- 2 1 1 出力指定情報
- 3 2 1 生物情報データベースクエリリスト
- 3 2 2 ユーザクエリ情報
- 3 3 1 生物情報データベースクエリ
- 4 1 1 第二の出力
- 5 1 1 文献検索問合せ
- 5 1 2 文献データベース検索結果
- 5 1 3 文献検索結果
- 6 1 1 文献検索解析結果
- 7 1 1 整形出力結果
- 9 9 9 生物情報データベース
- 1 0 0 0 文献データベース

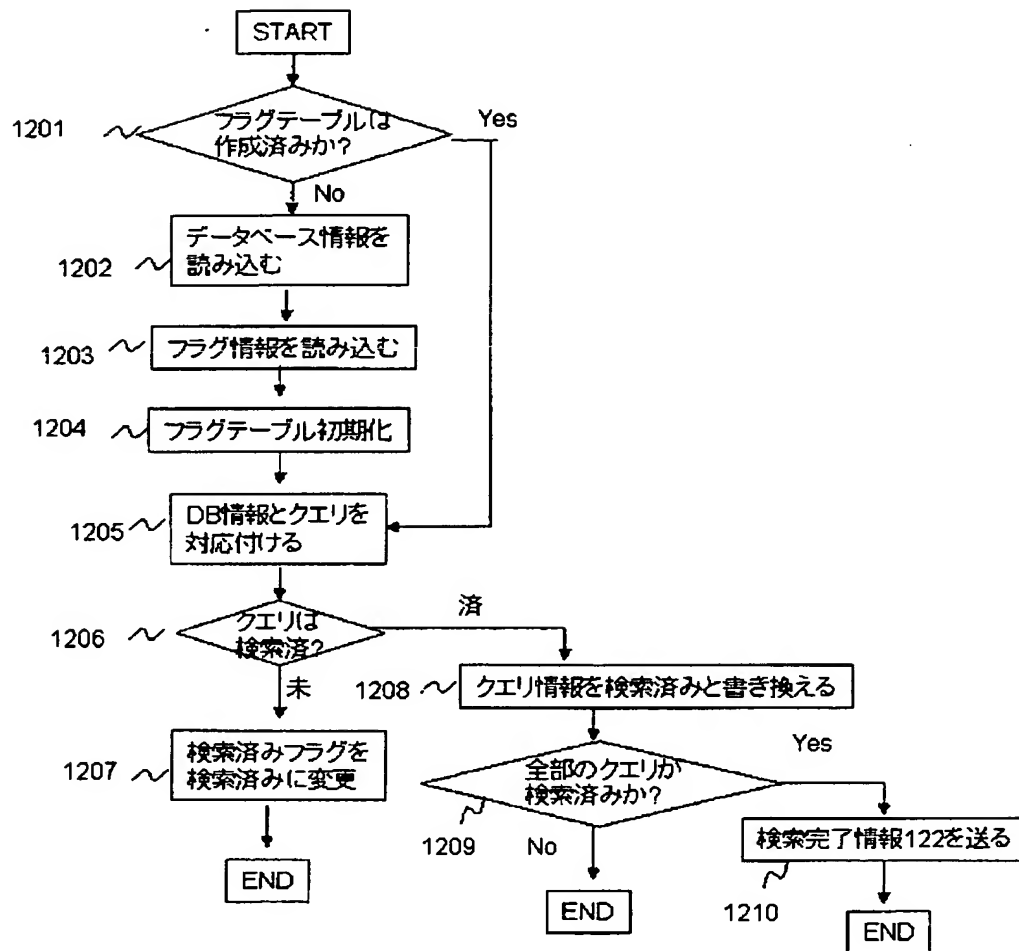




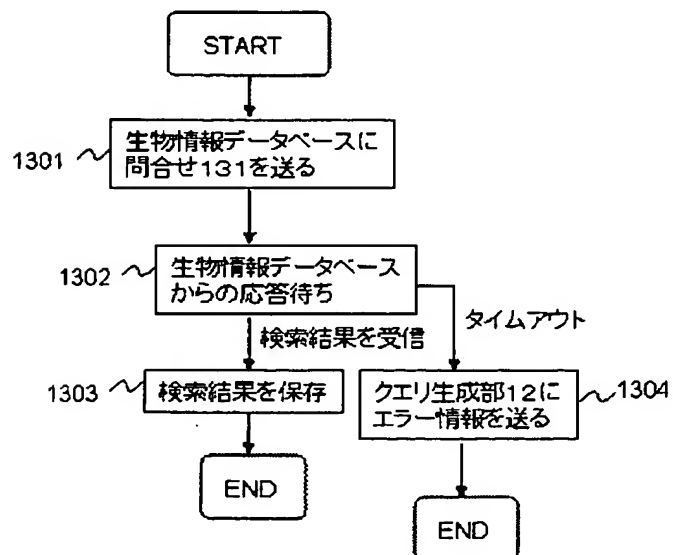
【図 2】



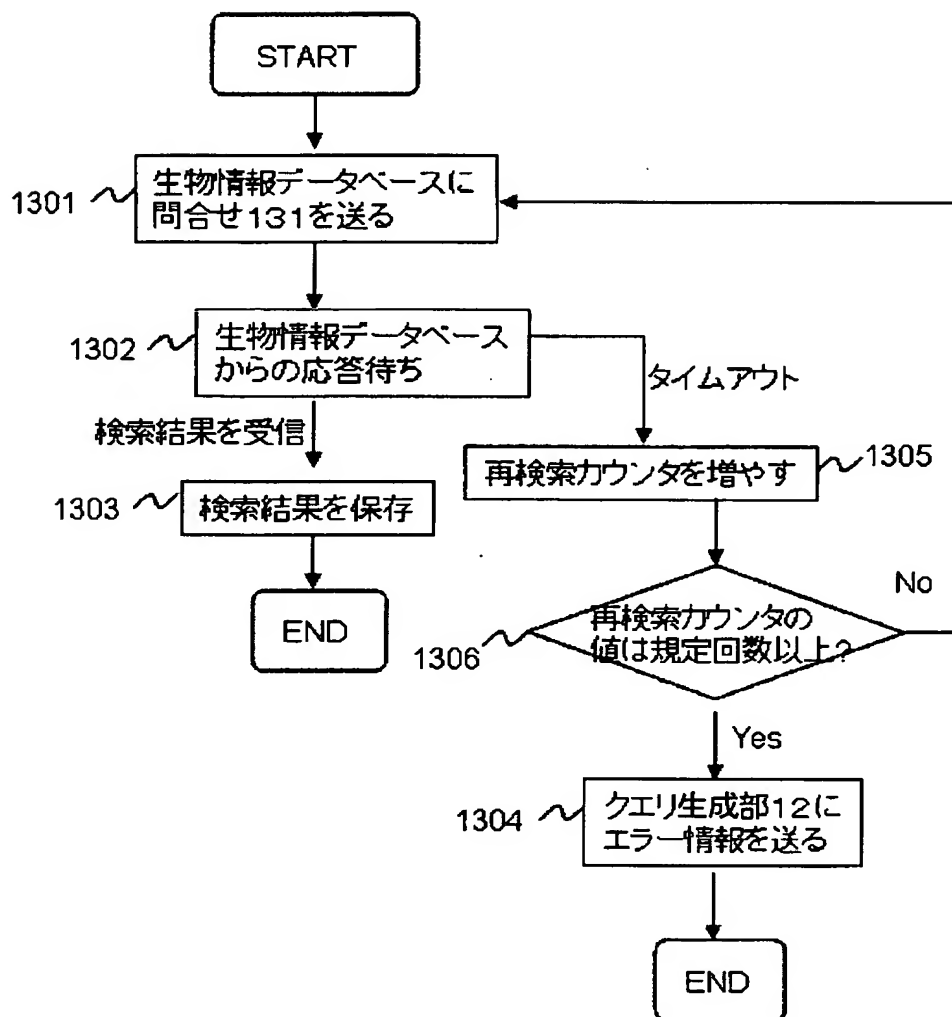
【図 3】



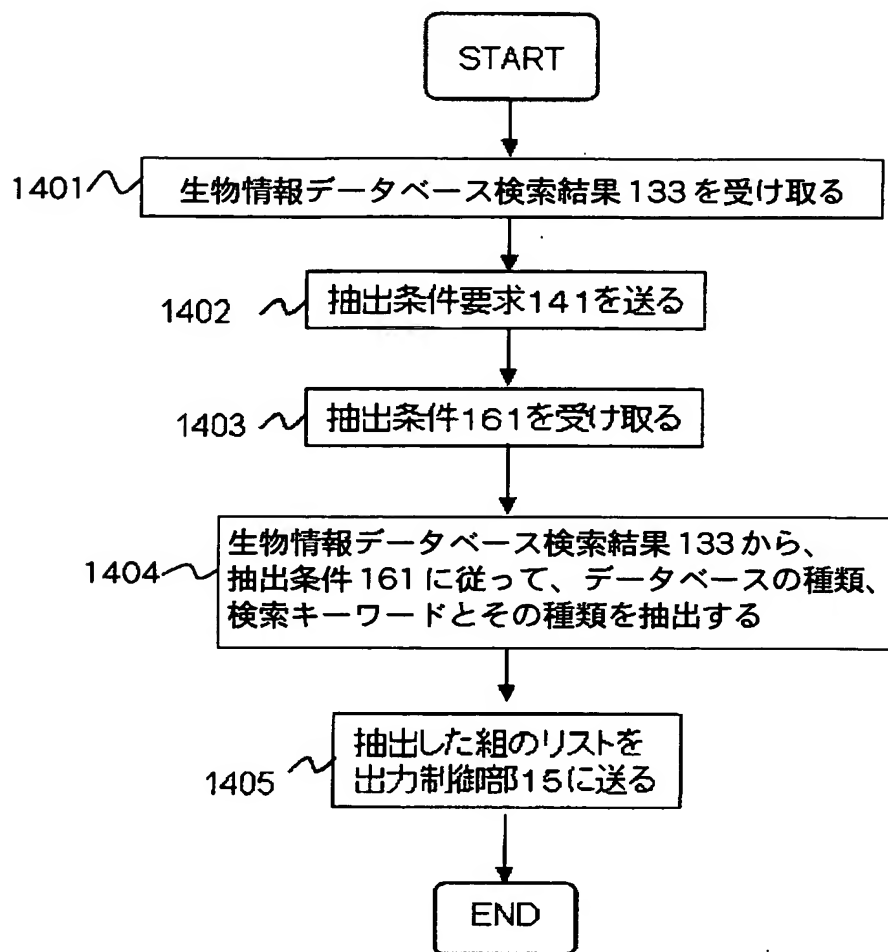
【図 4】



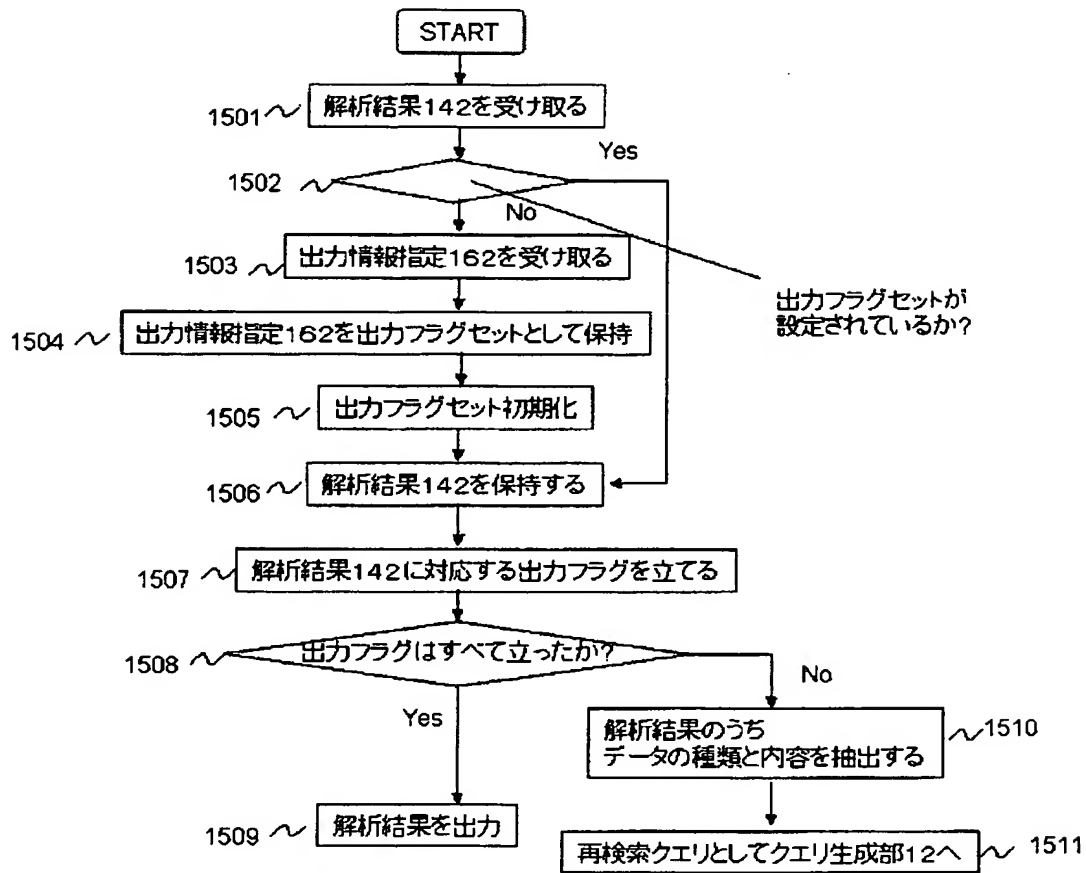
【図 5】



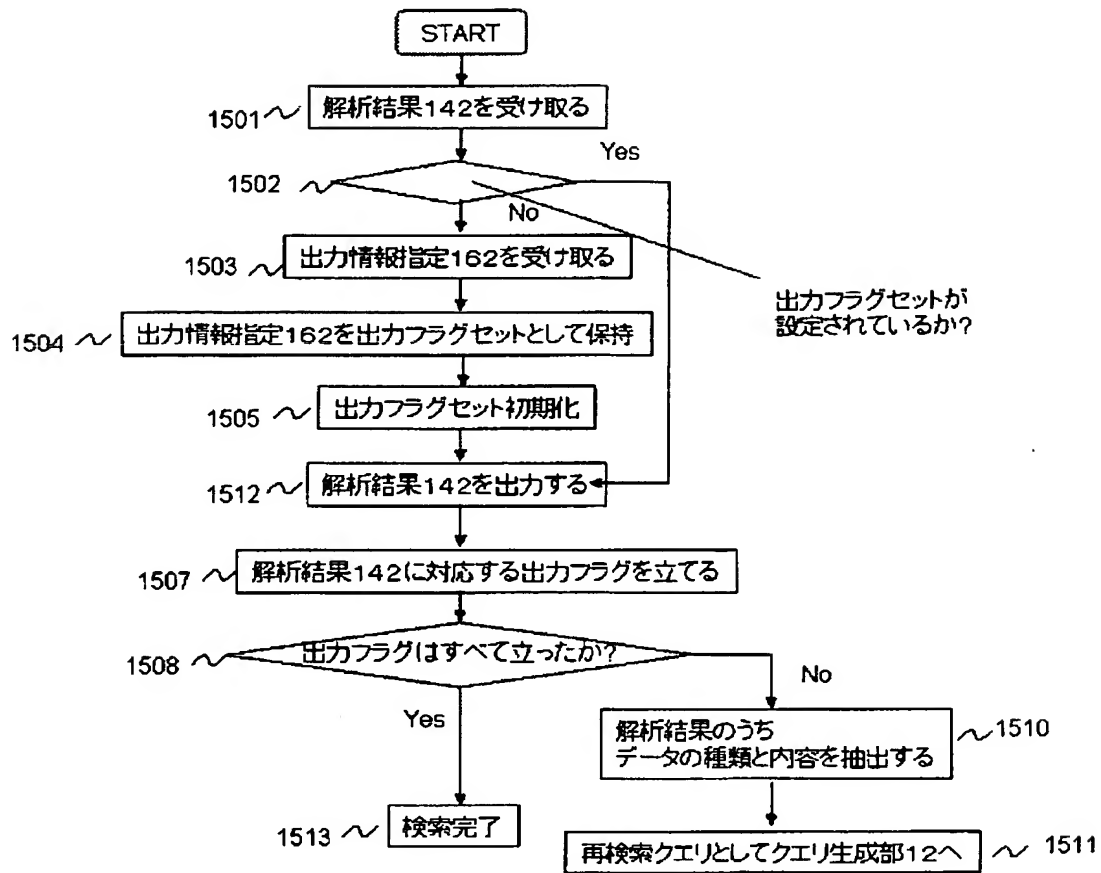
【図 6】



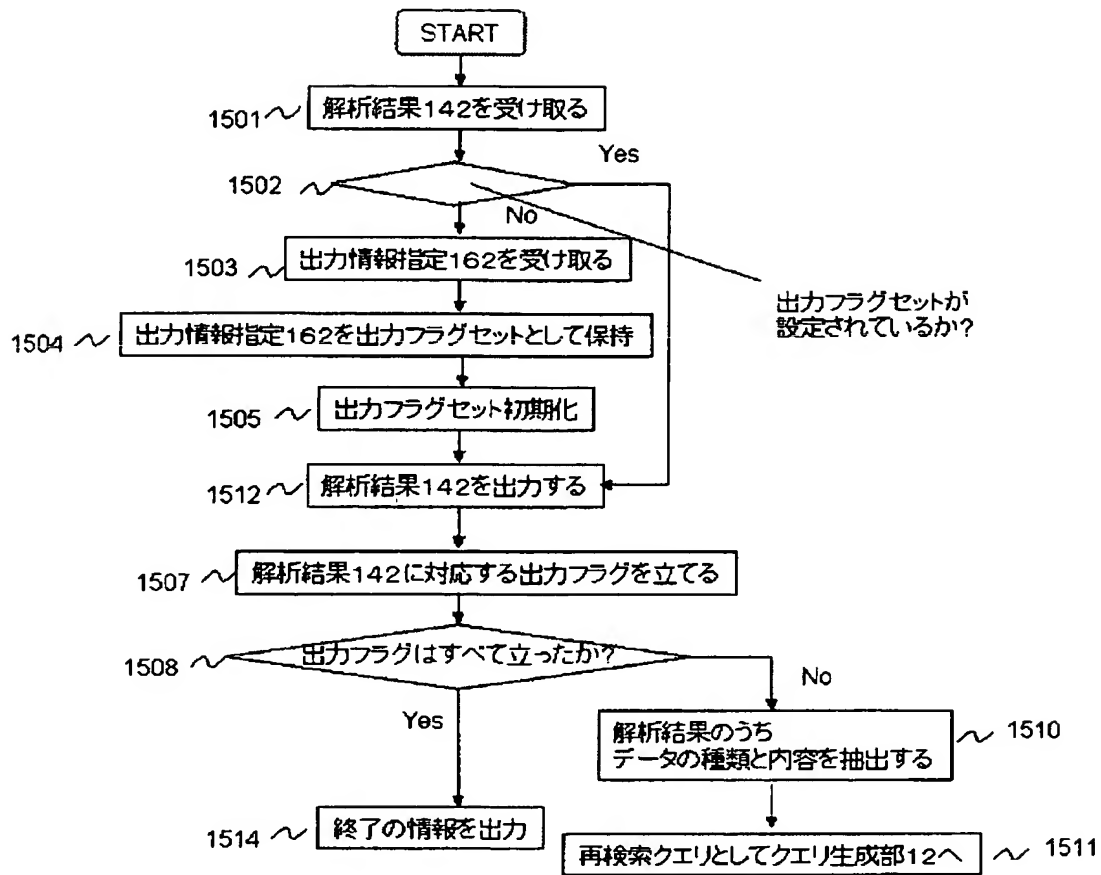
【図 7】



【図 8】

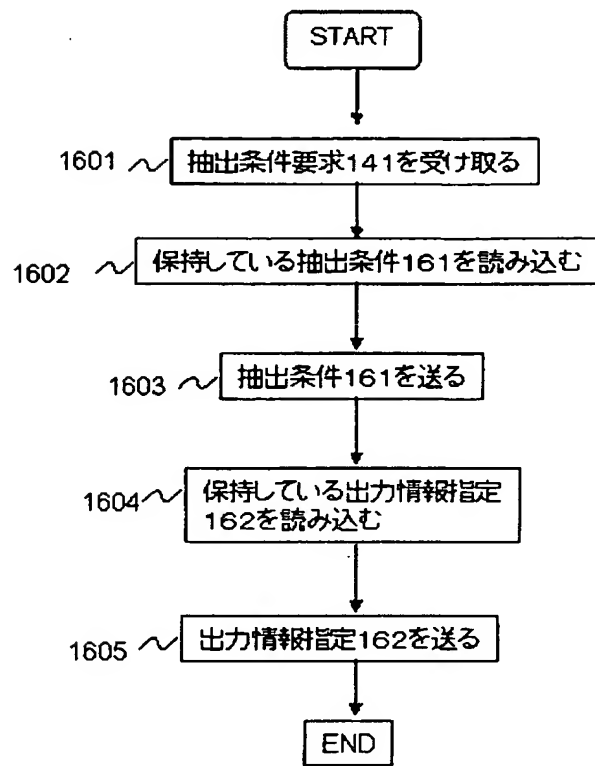


【図 9】

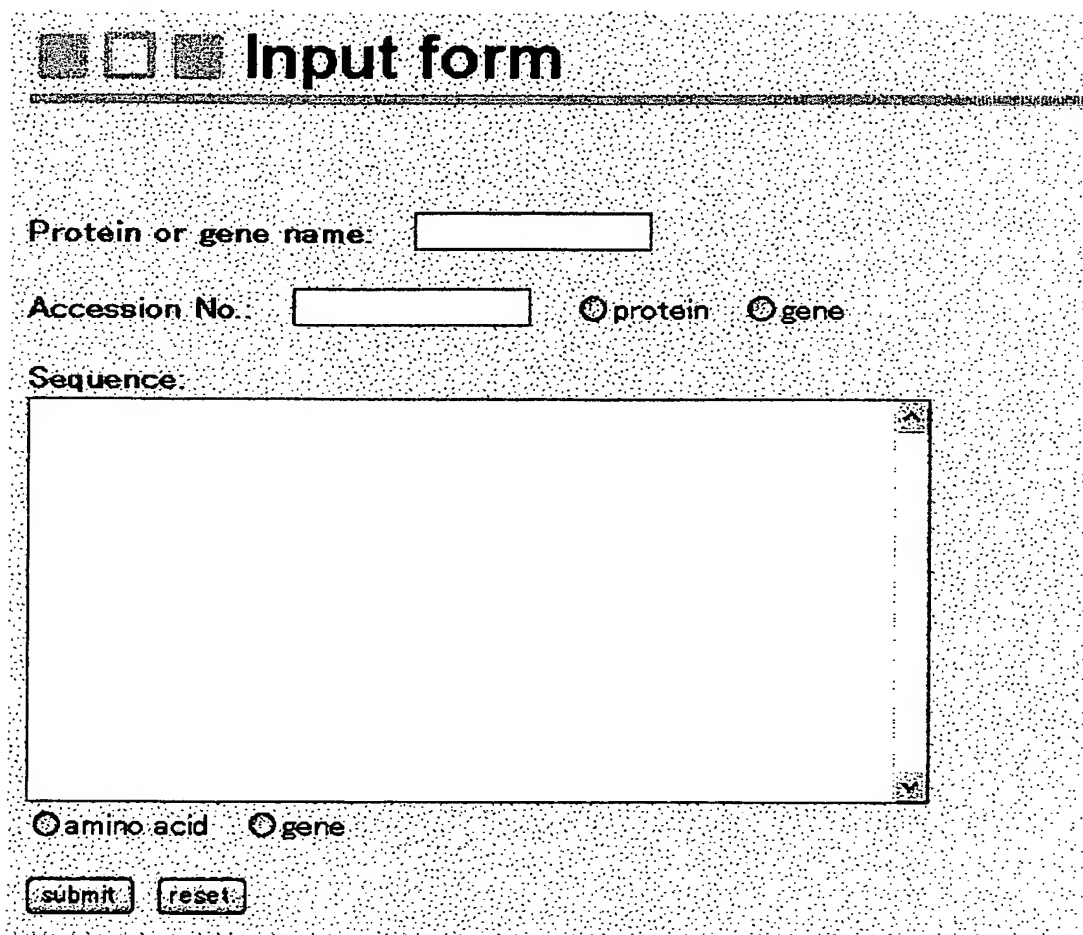




【図 10】



【図 11】



**Input form**

Protein or gene name:

Accession No.  ☒ protein ☐ gene

Sequence:

☒ amino acid ☐ gene

【図 12】

5172	Hs	SLC28A4	solute carrier family 26, member 4	7q31	100000
			Aliases:	PDS, DFNB4, PENDRIN	
			OMM:	605648	
別名として抽出する			RefSeq[R]:	NM_000441	
			Nucleotide:	AC002467, AF030880	
			Protein:	AAB88773, AAC51873, NP_000432	
23985	Mm	Slc28a4	solute carrier family 26, member 4	12 B1	100000
			Aliases:	Pds	
			RefSeq:	NM_011867	
			Nucleotide:	AF167411	
			Protein:	AAD51617, NP_035997	
29440	Rn	Slc28a4	solute carrier family 26, member 4	6q16	100000
			Aliases:	Pds	
			RefSeq:	NM_019214	
			Nucleotide:	AF167412	
			Protein:	AAD51618, NP_062087	
65010	Hs	SLC26A6	solute carrier family 26, member 6	3p21.3	100000
			Aliases:	DKFZp586E1422	
			RefSeq[R]:	NM_022911, NM_134263, NM_134426	
			Nucleotide:	AB033288, AB102713, AF279265, AF288410, AF416721, AK056237, AL050170, BC017697	
			Protein:	AAF81811, AAH17697, AAK19153, AAN07094, BAB69041, BAB71126, BAC56881, CAB43306, NP_075062, NP_599025, NP_602298	

【図 13】

Links

□1: NT\_039548  
Mus musculus chromosome 12 genomic contig, strain C57BL/6J  
gi|28520880|ref|NT\_039548.1|Mm12.39588.30[28520880]

Links

□2: NW\_043944  
Rattus norvegicus chromosome 6 WGS supercontig  
gi|26007775|ref|NW\_043944.1|Rn6.1303[26007775]

Links

□3: NM\_134426  
Homo sapiens solute carrier family 28, member 6 (SLC26A6), transcript variant 3, mRNA  
gi|20336278|ref|NM\_134426.1|[20336278]

⋮

Links

□35: AI916693  
tu89g1.1.x1 NCICGAP\_Gas4 Homo sapiens cDNA clone IMAGE:2258276 3' similar to TR043511 O43511 PENDRIN ;  
MRNA sequence  
gi|5636553|gb|AI916898.1|[5636553]

Links

□36: A747481  
u15h05.x1 Sugano mouse embryo mewa Mus musculus cDNA clone IMAGE2076921 3' similar to TR043511 O43511  
PENDRIN ; MRNA sequence  
gi|5125725|gb|A747481.1|[5125725]

Links

□37: AF030880  
Homo sapiens pendrin (PDS) mRNA, complete cds  
gi|2654004|gb|AF030880.1|AF030880[2654004]

アクセス番号      GI番号

## 【図 14】

LOCUS AF030880 4930 bp mRNA linear PRI 01-DEC-1997  
 DEFINITION Homo sapiens pendrin (PDS) mRNA, complete cds.  
 ACCESSION AF030880  
 VERSION AF030880.1 GI:2654004  
 KEYWORDS .  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 4930)  
 AUTHORS Everett,L.A., Glaser,B., Beck,J.C., Idol,J.R., Buchs,A., Heyman,M.,  
 Adawi,F., Hazani,E., Nassir,E., Baxevaris,A.D., Sheffield,V.S. and  
 Green,E.D.  
 TITLE Pendred syndrome is caused by mutations in a putative sulphate  
 transporter gene (PDS)  
 JOURNAL Nat. Genet. 17 (4), 411-422 (1997)  
 MEDLINE 98081089  
 PUBMED 9838842  
 REFERENCE 2 (bases 1 to 4930)  
 AUTHORS Everett,L.A., Glaser,B., Beck,J.C., Idol,J.R., Buchs,A., Heyman,M.,  
 Adawi,F., Hazani,E., Nassir,E., Baxevaris,A.D., Sheffield,V.S. and  
 Green,E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-OCT-1997) Genome Technology Branch, National Human  
 Genome Research Institute, National Institutes of Health, 49  
 Convent Drive, Bethesda, MD 20892, USA

FEATURES  
 source Location/Qualifiers  
 1..4930  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q22-q31.1"  
 gene 1..4930  
 /gene="PDS" 別名として抽出する  
 CDS 225..2587  
 /gene="PDS"  
 /function="putative sulfate transporter"  
 /note="mutated in Pendred syndrome"  
 /codon\_start=1  
 /product="pendrin"  
 /protein\_id="AACST979.1"  
 /db\_xref="GI:2654005"  
 /translation="MAAPGRSEPPOLPEYSCSYNWSRPVYSELAFDQGHERRLOERK  
 TLRESLAKCCSCSRKRAFGVLKTLVPILMLPKYRKERLLSDVISGVSTQLVATLOG  
 :  
 EELDIPKTEIEIQVDINSELPVKVMFKVPILSLVLDGAI SFLDWGVRSRLVIVKE  
 FORIDVNVYFASLDQYVIEKLEOCGFDDNIRKDTFFLTVDAILYLQNVKSGEGG  
 SILETITLIDCKDTLELLETETEEFLDWDEAMRTLAS" アミノ酸配列として抽出する

BASE COUNT 1454 a 937 c 1082 g 1457 t

## ORIGIN

```

1 ctcagccttc ccgcttcagg aaazgggagz aatgcaggaz ggtazgati tcttctga
61 taggatcgtt tggaaagac cgcagcctgt gtgtgtcttt cocttcgacc aazgtatctg
121 ttgtctcgta aataaaacgt cccactgocf tctgagagcz ctataaagac aazgzaagaz
:
4301 tccagtattg tatatgasti ttaacaaatt aaaaaatcaa atcatgtaca ttgaaaaata
4361 ttgtcacaca tttaaaaata aatgtaaagt tgtcttttaa actactcgga tgtgtccttt
4921 ctgaacaaaa

```

Revised: July 5, 2002.

塩基配列として抽出する

## 【図 15】

BLASTP 2.2.3 [Apr-24-2002]

Sequences producing significant alignments:	Score (bits)	E Value
ref NP_000432.1  pendrin [Homo sapiens] >gi 11421915 ref XP_0049...	1229	0.0
ref NP_062087.1  pendrin: Pendred syndrome homolog (human) [Ratt...	1101	0.0
ref NP_035997.1  pendrin: Pendred's syndrome [Mus musculus] >gi ...	1098	0.0
...		
pir  B82127 sulfate permease family protein VC2031 [imported] - ...	35	2.8
emb CAB62519.1  (AL050358) hypothetical protein [Homo sapiens]	34	3.6
pir  T16522 hypothetical protein K02E10.2 - Caenorhabditis elegans...	34	4.7

>ref|NP\_000432.1| **pendrin** [Homo sapiens]      名前として抽出する  
ref|XP\_004953.1| **pendrin** [Homo sapiens]  
sp|043511|PEND\_HUMAN PENDRIN (SODIUM-INDEPENDENT CHLORIDE/IODIDE TRANSPORTER)  
gb|AAC51873.1| (AF030880) pendrin [Homo sapiens]  
Length = 780

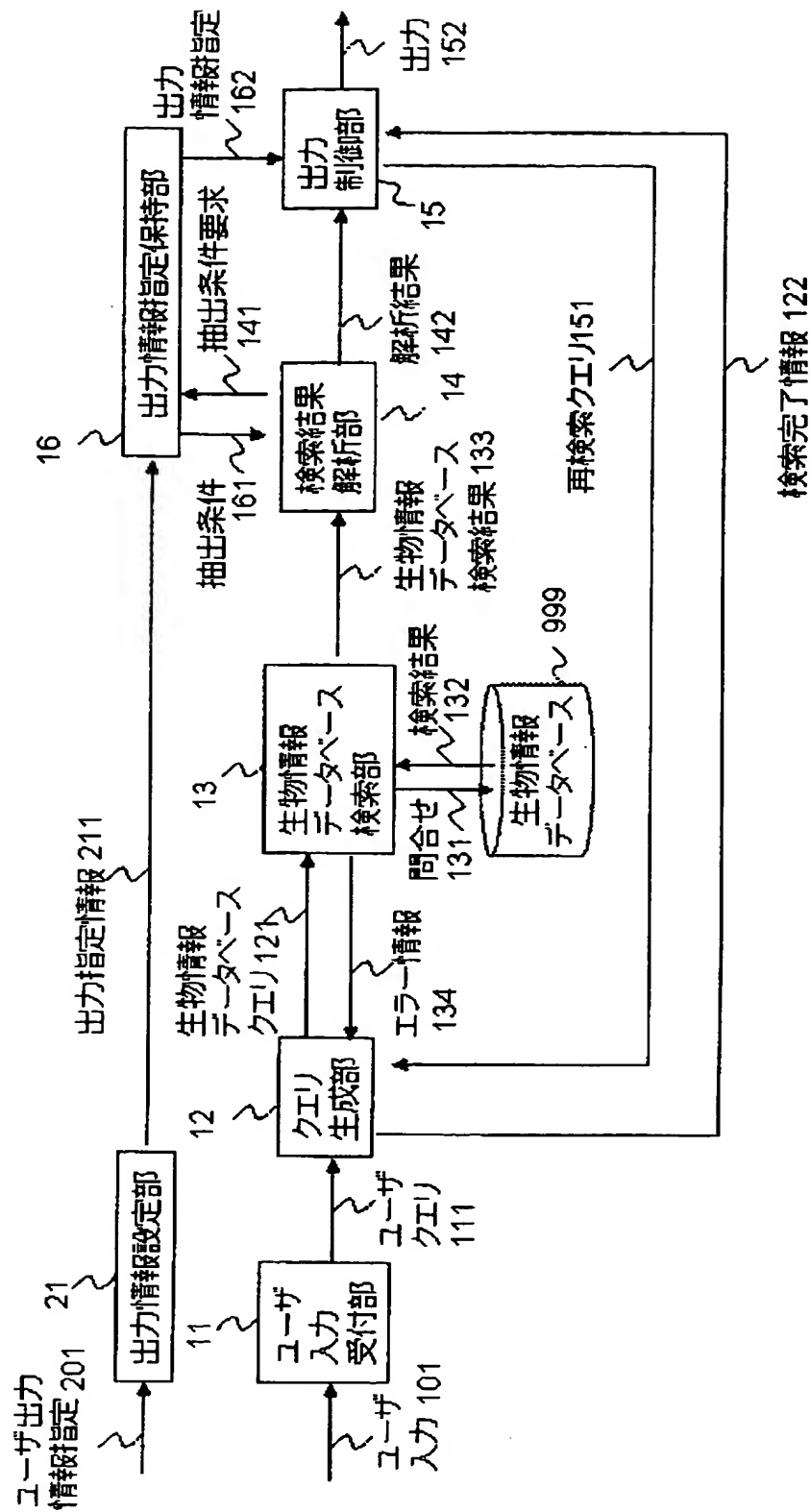
Score = 1229 bits (3180), Expect = 0.0  
Identities = 649/780 (83%), Positives = 649/780 (83%)

Query: 1    MAAPGGRSEPPQLPEYSCSYMVSRLPVYSELAFQOOHERRLQERKTLRESLAKCCSCSRKR 60  
          MAAPGGRSEPPQLPEYSCSYMVSRLPVYSELAFQOOHERRLQERKTLRESLAKCCSCSRKR  
Sbjct: 1    MAAPGGRSEPPQLPEYSCSYMVSRLPVYSELAFQOOHERRLQERKTLRESLAKCCSCSRKR 60

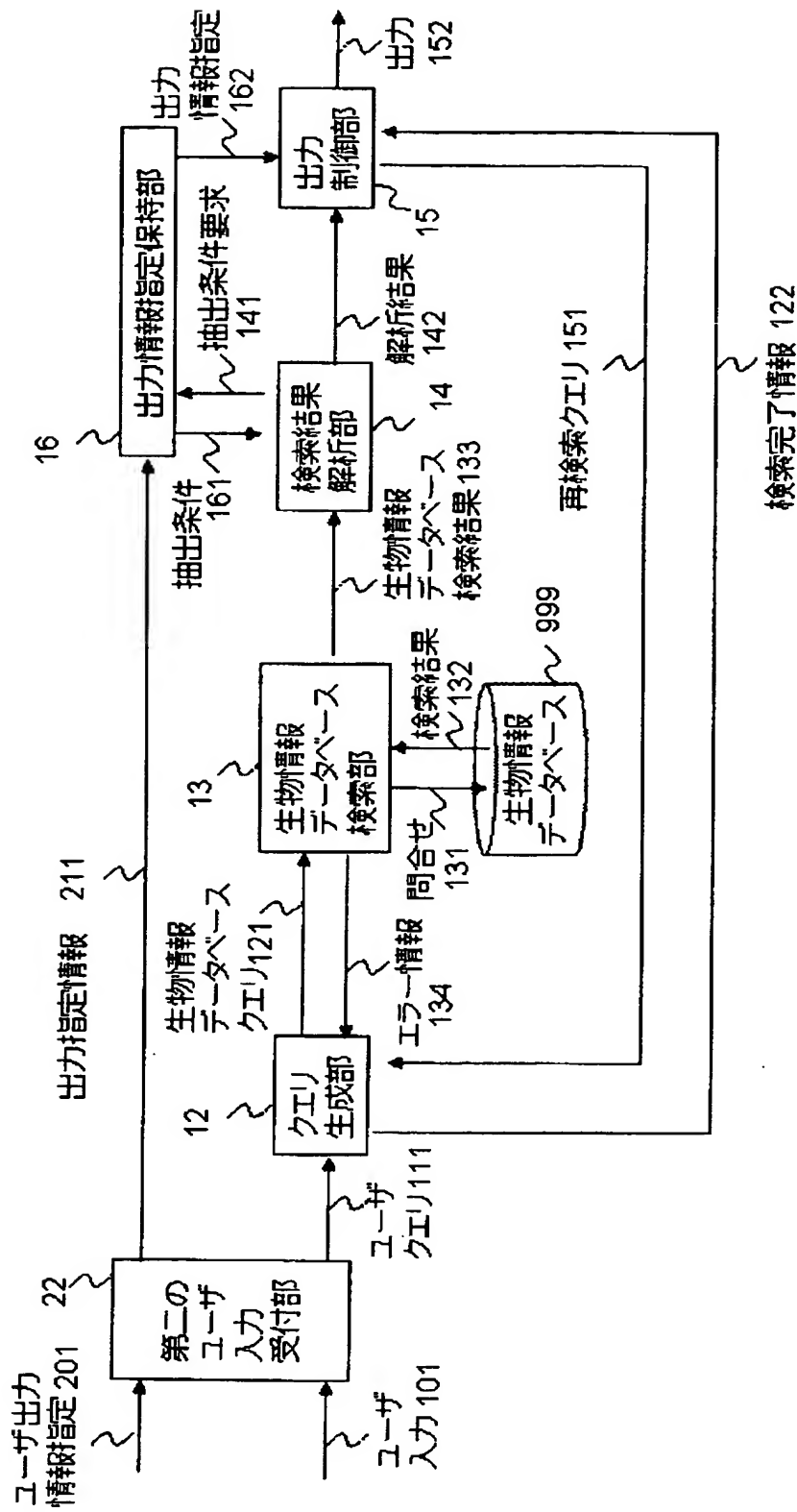
...

Query: 721 TVHDAILYLONOVKSQEGGGSILETITLIDCKDXXXXXXXXXXXXXXXXXDVQDEAMRTLAS 780  
          TVHDAILYLONOVKSQEGGGSILETITLIDCKD                    DVQDEAMRTLAS  
Sbjct: 721 TVHDAILYLONOVKSQEGGGSILETITLIDCKDTLEL IETELTEEELDVQDEAMRTLAS 780

【図 16】



【図 17】

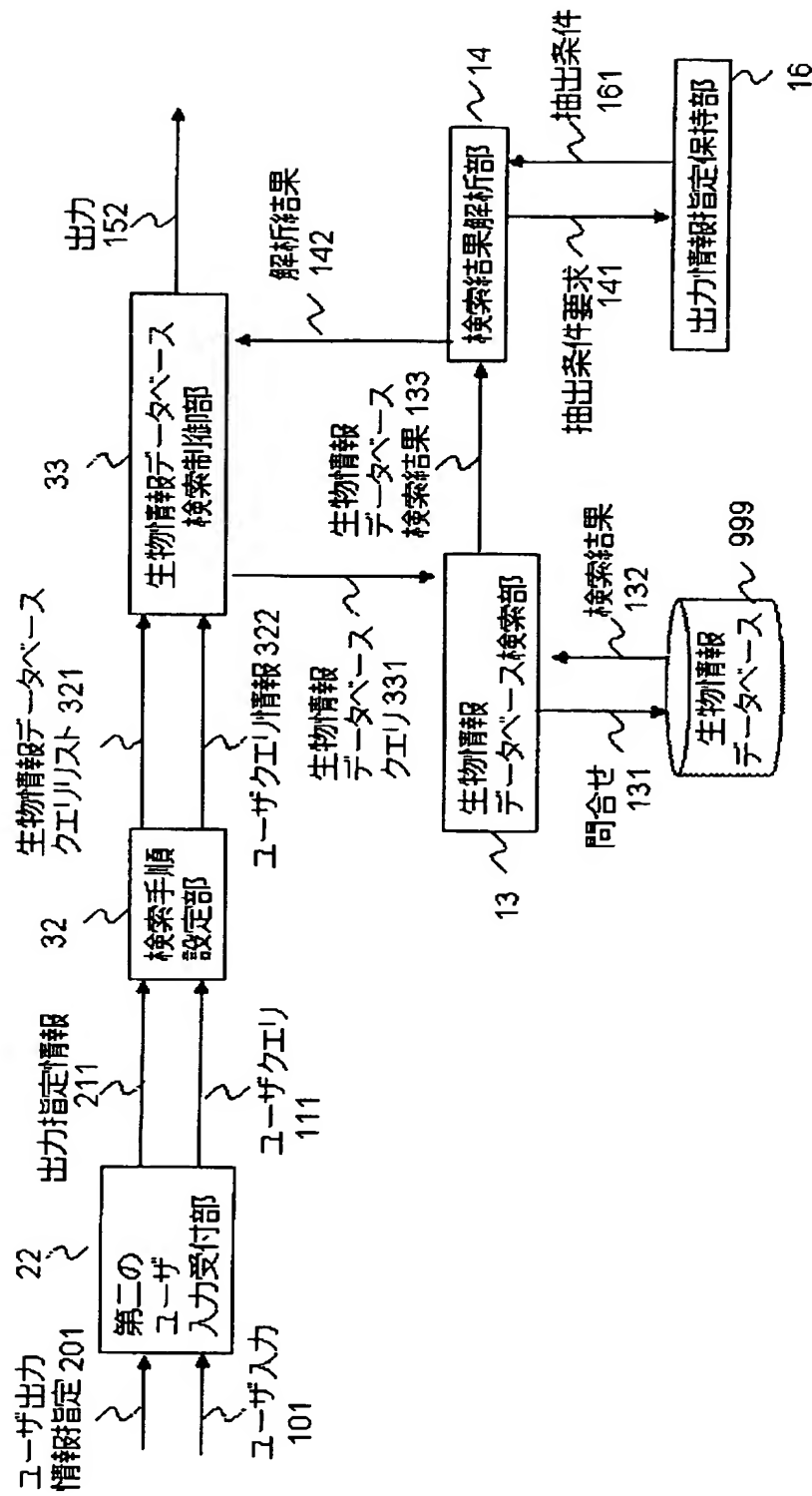




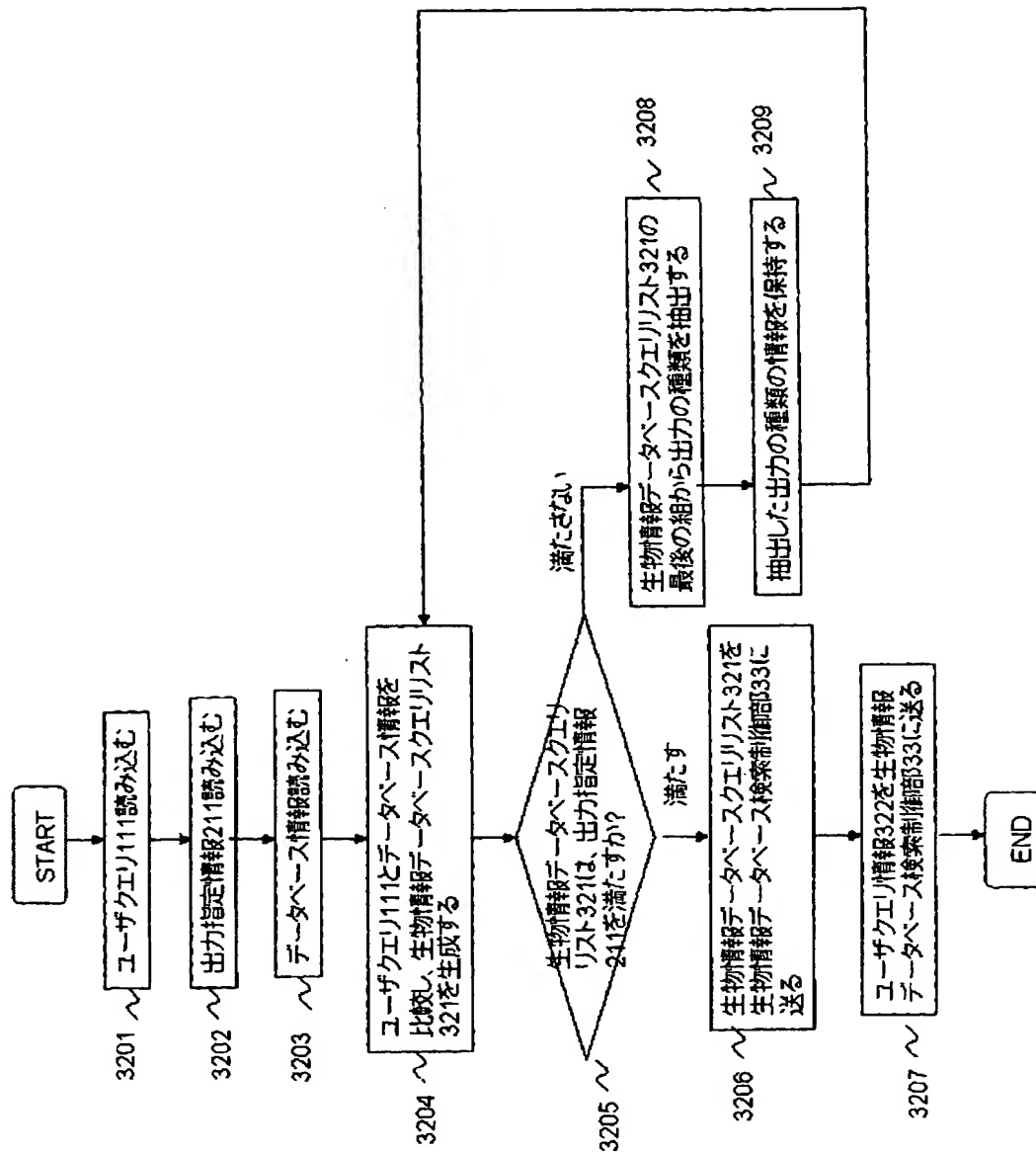
【図 18】

Entrez ☒ name ☐ Accession No. ☐ Sequence  
LocusLink ☒ name ☐ Accession No. (RefSeq) ☐ Accession No. (OMIM) ☐ Accession No. (Entrez) ☐ chromosome position  
BLAST ☒ name ☐ Accession No. ☐ Sequence --- e-value

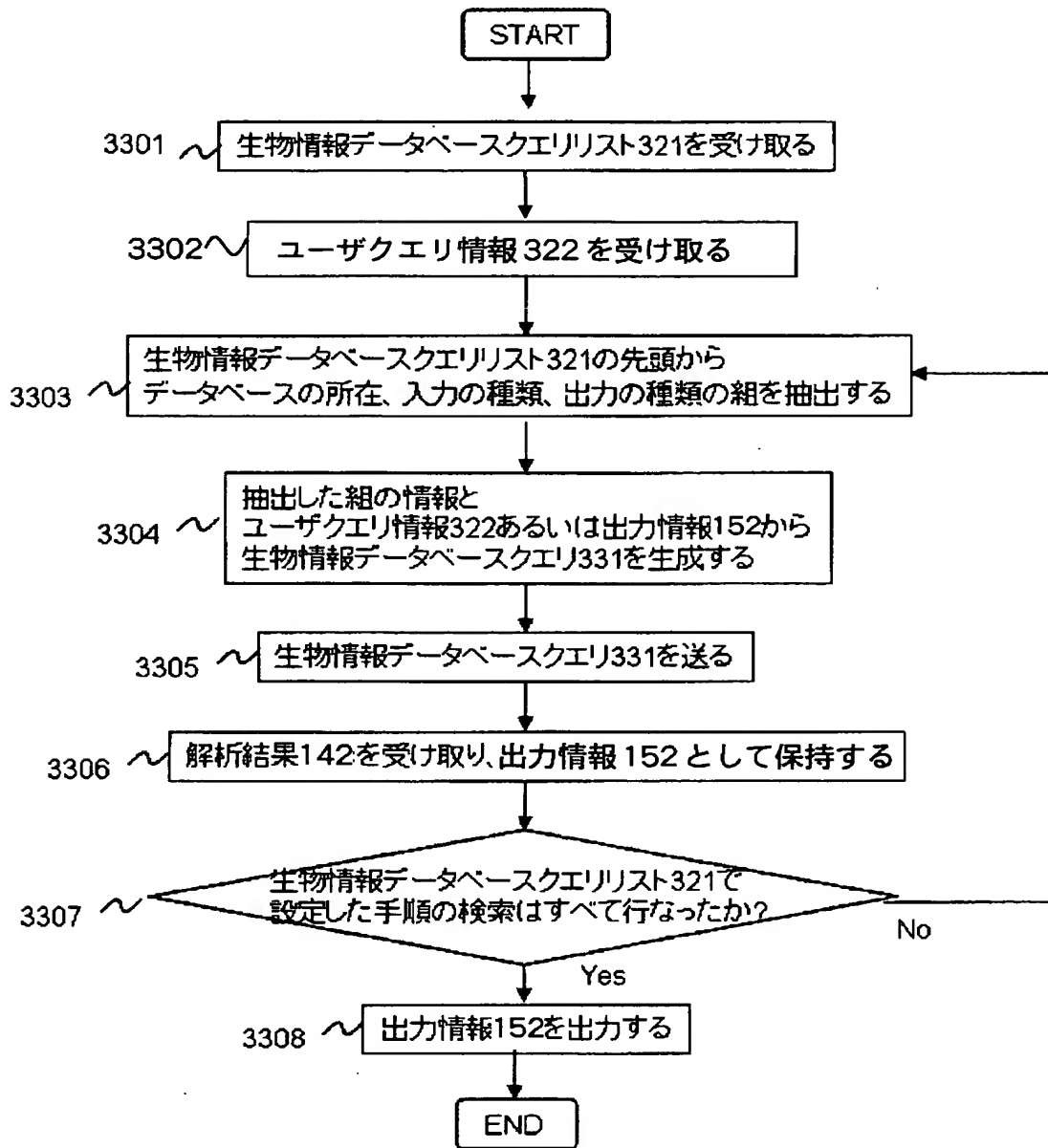
【図 19】



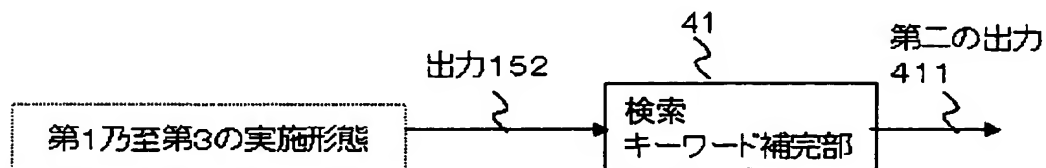
【図20】



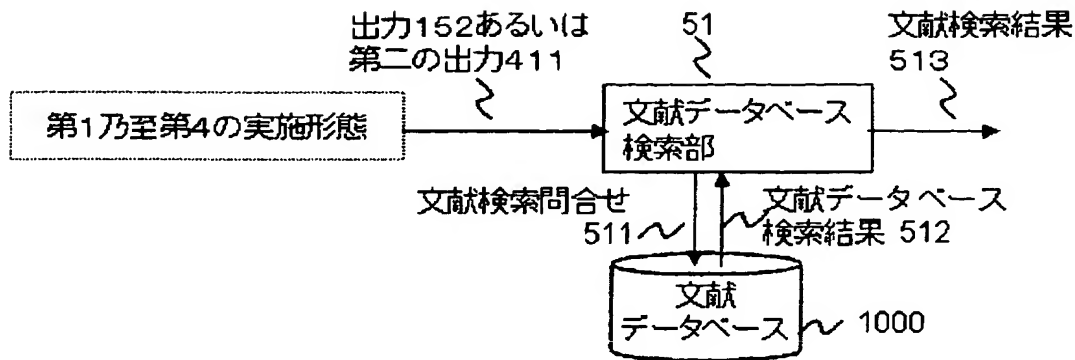
【図 2 1】



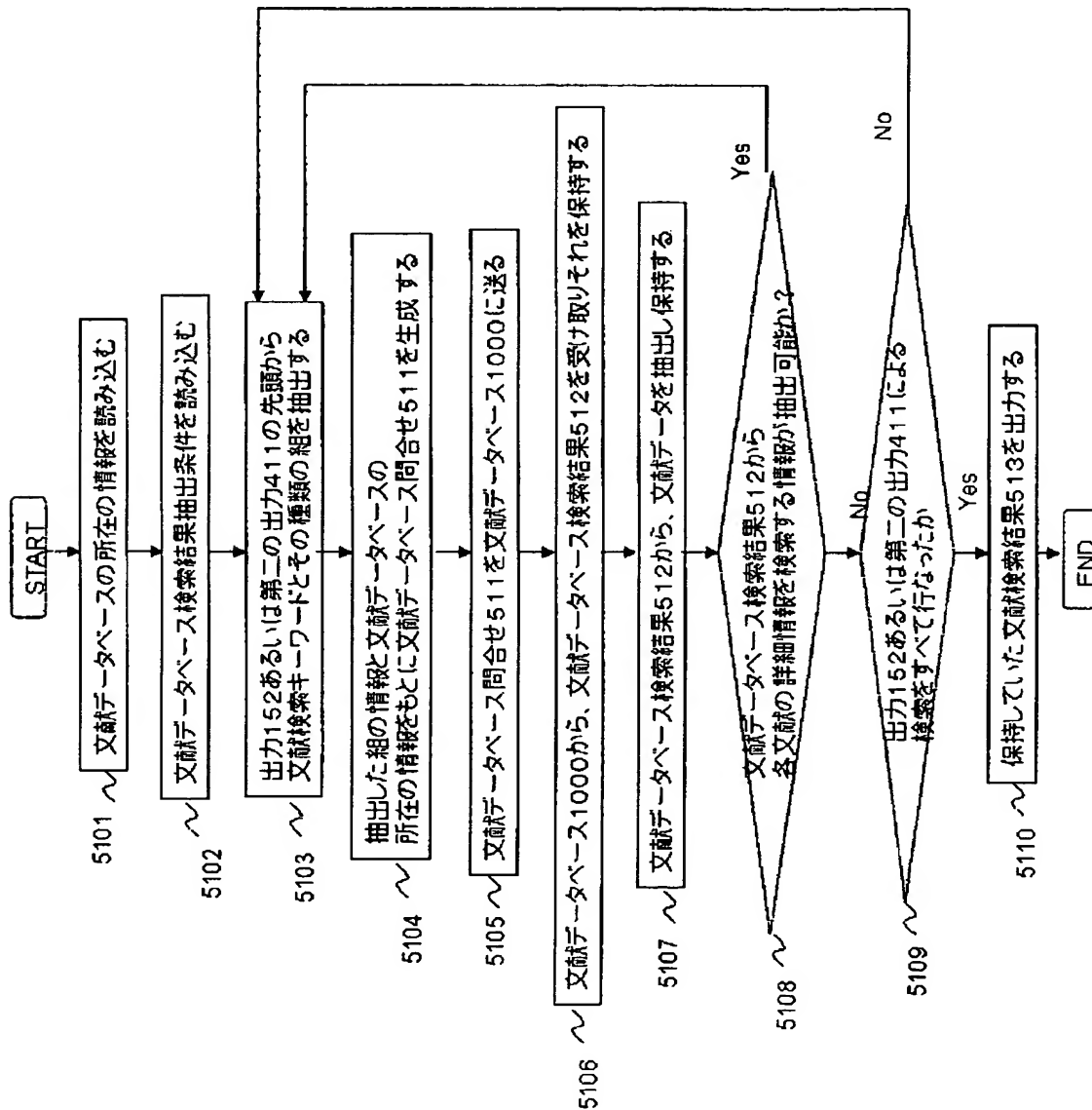
【図 2 2】



【図 2 3】



【図 2 4】



【図 25】

- 1: Huisander M, Kiernan AE, Ellingqvist SR, Carlsen P, Samuelsson E, Johansson BR, Steel KP, Enerback S. Lack of pendrin expression leads to deafness and expansion of the endolymphatic compartment in inner ears of Foxd1 null mutant mice. Development. 2003 May 1;130(9):2013-2026. PMID: 12642503 [PubMed - as supplied by publisher] Related Articles, Links
- 2: Gruters A, Biebermann H, Krude H. Neonatal thyroid disorders. Horm Res. 2003;59 Suppl 1:24-9. PMID: 12566717 [PubMed - in process] Related Articles, Links
- 3: Fricke S, Kwon TH, Frokjaer J, Madsen KM, Nielsen S. Regulated expression of pendrin in rat kidney in response to chronic NH4Cl or NaHCO3 loading. Am J Physiol Renal Physiol. 2003 Mar;284(3):F694-93. PMID: 12556366 [PubMed - indexed for MEDLINE] Related Articles, Links
- 4: Kondo T, Nakamura N, Suzuki K, Murata S, Muramatsu A, Kawasaki A, Katoh R. Expression of human pendrin in diseased thyroids. J Histochem Cytochem. 2003 Feb;51(2):167-73. PMID: 12535525 [PubMed - indexed for MEDLINE] Related Articles, Links
- 5: Zarnegar R, Eturied L, Kanauchi H, Wong M, Fung M, Ginzinger D, Duh QY, Clark OH. Increasing the effectiveness of radioactive iodine therapy in the treatment of thyroid cancer using Trichostatin A, a histone deacetylase inhibitor. Surgery. 2002 Dec;132(6):984-90; discussion 990. PMID: 12490845 [PubMed - indexed for MEDLINE] Related Articles, Links
- 6: Wagner CA, Fintberg KE, Stehberger FA, Litton RP, Giebisch GH, Aronson PS, Gallei JP. Regulation of the expression of the Cl-/anion exchanger pendrin in mouse kidney by acid-base status. Kidney Int. 2002 Dec;62(6):2109-17. PMID: 12427135 [PubMed - in process] Related Articles, Links
- 7: Ko SE, Shechevnikov N, Choi JY, Luo X, Ishibashi K, Thomas P, Kim JY, Kim KH, Lee MG, Naruse S, Musilem S. A molecular mechanism for aberrant CFTR-dependent HCO(3X-) transport in cystic fibrosis. EMBO J. 2002 Nov 1;21(21):5662-72. PMID: 12411484 [PubMed - indexed for MEDLINE] Related Articles, Links

1: Development 2003 May 1;130(9):2013-2025



Related Articles, Links

【図 26】

# **Lack of pendrin expression leads to deafness and expansion of the endolymphatic compartment in inner ears of Foxi1 null mutant mice.**

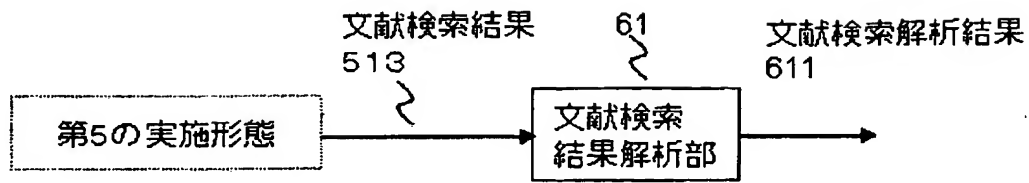
Hulander M, Kiernan AE, Blomqvist SR, Carlsson P, Samuelsson EJ, Johansson BR, Steel KP, Enerback S.

Medical Genetics, Department of Medical Biochemistry, Institute of Anatomy and Cell Biology, Goteborg University, Box 440, SE-405 30 Goteborg, Sweden. MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK. The Electron Microscopy Unit, Institute of Anatomy and Cell Biology, Goteborg University, Box 440, SE-405 30 Goteborg, Sweden. Department of Molecular Biology, Institute of Anatomy and Cell Biology, Goteborg University, Box 440, SE-405 30 Goteborg, Sweden.

Mice that lack the winged helix/forkhead gene Foxi1 (also known as Fkh10) are deaf and display shaker/waltzer behavior, an indication of disturbed balance. While Foxi1 is expressed in the entire otic vesicle at E9.5, it becomes gradually restricted to the endolymphatic duct/sac epithelium and at E16.5 Foxi1 expression in the inner ear is confined to this epithelium. Histological sections, paintfill experiments and whole-mount hybridizations reveal no abnormality in inner ear development of Foxi1(-/-) mice before E13.5. Between E13.5 and E16.5 the membranous labyrinth of inner ears from null mutants starts to expand as can be seen in histological sections, paint-fill experiments and three-dimensional reconstruction. Postnatally, inner ears of Foxi1(-/-) mice are extremely expanded, and large irregular cavities, compressing the cerebellum and the otherwise normal middle ear, have replaced the delicate compartments of the wild-type inner ear. This phenotype resembles that of the human sensorineural deafness syndrome Pendred syndrome, caused by mutations in the PDS gene. In situ hybridization of Foxi1(-/-) endolymphatic duct/sac epithelium shows a complete lack of the transcript encoding the chloride/iodide transporter pendrin. Based on this, we would like to suggest that Foxi1 is an upstream regulator of pendrin and that the phenotype seen in Foxi1 null mice is, at least in part, due to defective pendrin-mediated chloride ion resorption in the endolymphatic duct/sac epithelium. We show that this regulation could be mediated by absence of a specific endolymphatic cell type - FORE (forkhead related) cells - expressing Foxi1, Pds, Coch and Jag1. Thus, mutations in FOXI1 could prove to cause a Pendred syndrome-like human deafness.

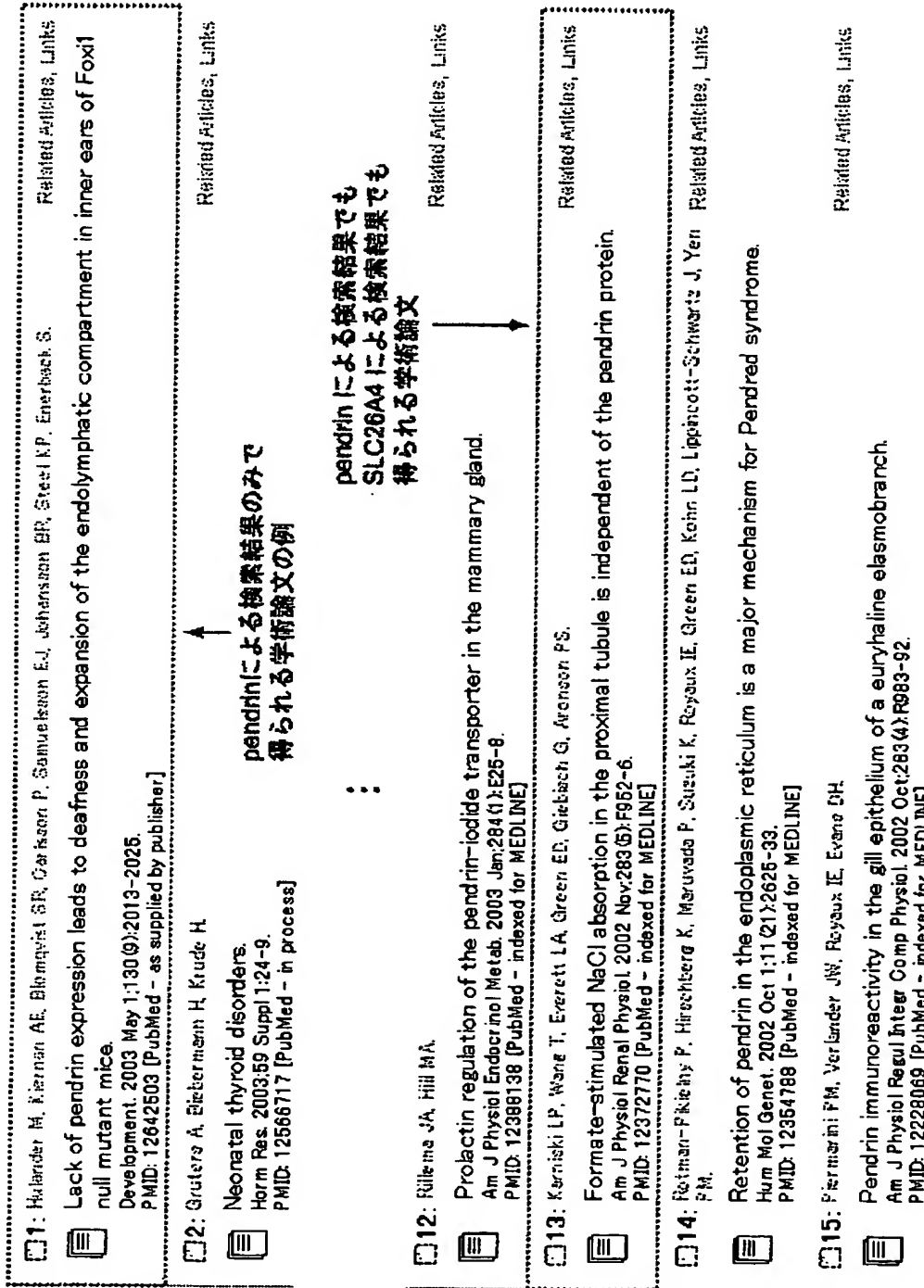
PMID: 12642503 [PubMed - as supplied by publisher]

【図 27】





【図 28】



【図 29】

pendrin による検索結果でも  
SLC26A4 による検索結果でも  
得られる学術論文



	Related Articles, Links
[1] Karimski LP, Ward J, Everett LA, Green ED, Giebisch G, Aronson PS. Formate-stimulated NaCl absorption in the proximal tubule is independent of the pendrin protein. Am J Physiol Renal Physiol. 2002 Nov;283(5):F952-6. PMID: 12372770 [PubMed - indexed for MEDLINE]	
[2] Blaser-Gondky M. Hereditary deafness and phenotyping in humans. Br Med Bull. 2002;59:73-94. Review. PMID: 12324385 [PubMed - indexed for MEDLINE]	

[3] Smith RJ, Robin NH

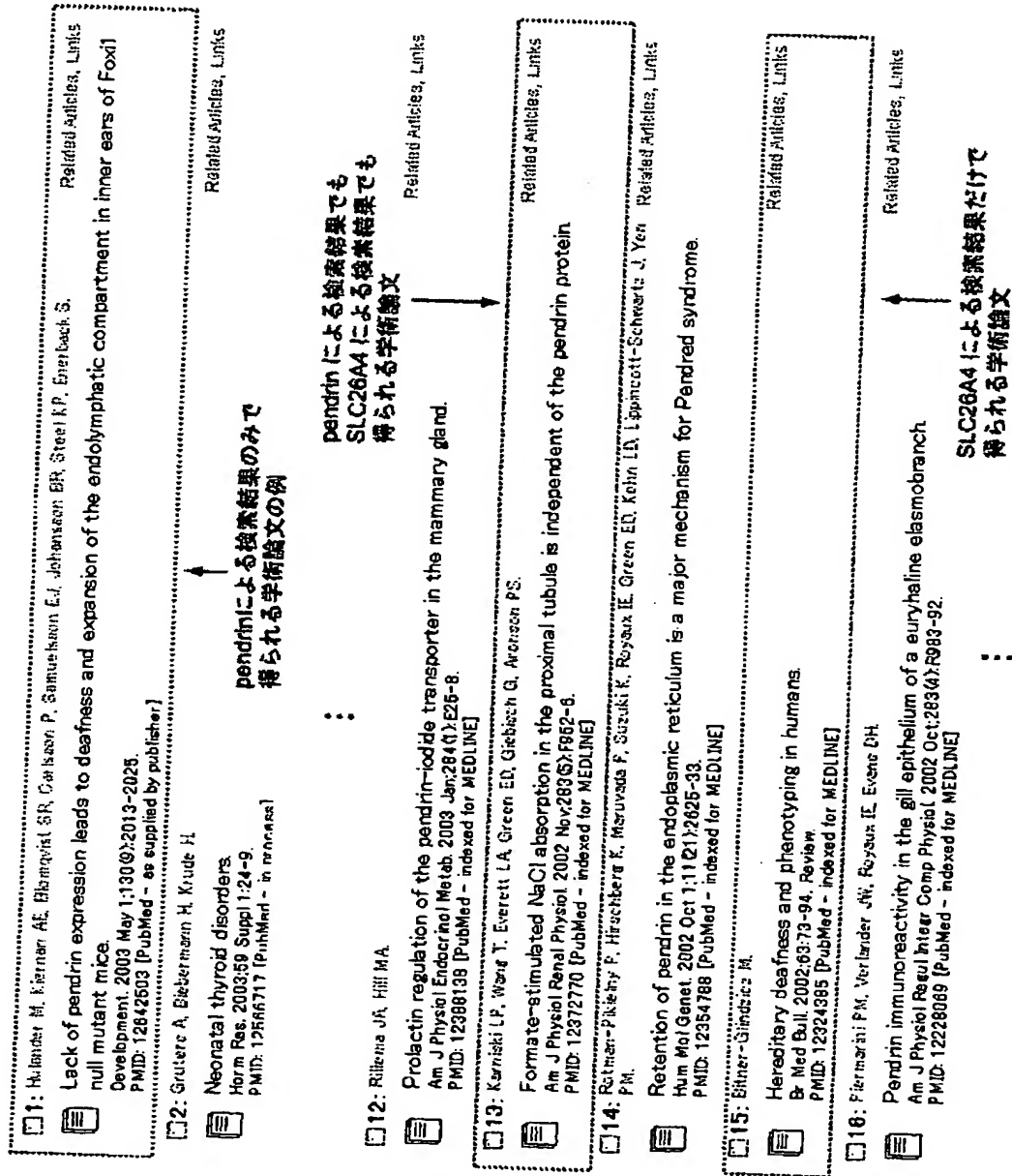
Genetic testing for deafness--GJB2 and SLC26A4 as causes of deafness.  
J Commun Disord. 2002 Jul-Aug;35(4):367-77.  
PMID: 12160355 [PubMed - indexed for MEDLINE]

⋮

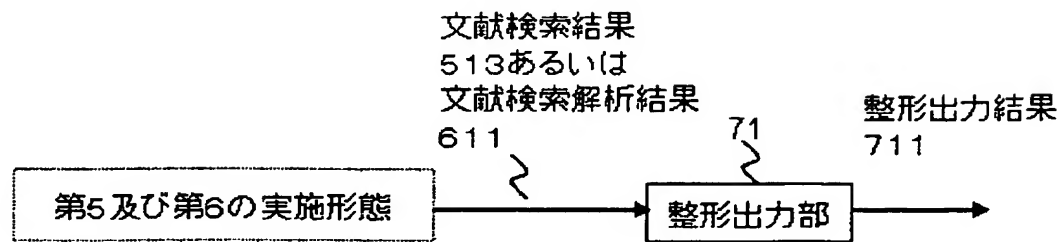
SLC26A4 による検索結果だけで  
得られる学術論文



【図 30】



【図 3 1】



【図32】

Related Articles, Links

Thyroid 2001 Nov;11(11):1017-23

# Expression of nicotinamide adenine dinucleotide phosphate oxidase flavoprotein DUOX genes and proteins in human papillary and follicular thyroid carcinomas.

Lacroix L, Nocera M, Mian C, Gaillou B, Viron A, Dupuy C, Filletti S, Bidart JM, Schlumberger M

Department of Clinical Biology, Institut Gustave-Roussy, Villejuif, France.

Duox2, and probably Duox1 are glycoproteins involved in the thyroid H2O2 generator functionsally associated to thyroperoxidases (TPO). We investigated both DUOX1 and DUOX2 gene expressions using quantitative reverse transcription-polymerase chain reaction (RT-PCR) in 47 thyroid carcinomas, including 10 paired normal/tumoral tissues. In carcinomas, variations of DUOX1 and DUOX2 mRNA levels were parallel, indicating that control mechanisms of both gene expressions operate in tumors as well as in normal thyroid tissues; DUOX1 expression was in the normal range in 20, was decreased up to 50-fold in 8, and increased up to 7-fold in 19 samples. DUOX2 expression was in the normal range in 15, was decreased up to 200-fold in 10, and increased up to 5-fold in 22 samples. In the 10 paired samples, variations of DUOX and TPO gene expressions were not correlated. We analyzed Duox2/2 protein expression in 88 tumor samples using an antipeptide antisera reacting with both Duox proteins. In normal tissues, Duox proteins are localized at the apical pole of thyrocytes, with 40% to 60% of thyrocytes being stained. In the 88 cancer tissues, immunostaining was absent in 19 samples, was low in 32, and normal or even slightly increased in the other 35 samples. The expression of Duox proteins was related to tumor differentiation, being more frequently found in neoplastic tissues that were able to pick up radioiodine, and in those with a detectable expression of sodium iodide symporter (NIS), pendrin and TPO.

PMD: 11782710 [PubMed - indexed for MEDLINE]

文献の概要をリスト形式で表示

Thyroid 2001 Nov;11(11):1017-23

キーワードの表示色を変える

Related Articles, Links

## 遺伝子あるいはタンパク質の別名を表示 Update on intrathyroidal iodine metabolism.

Dunn JT, Dunn AD.

Department of Medicine, University of Virginia, Charlottesville, USA. jtd@virginia.edu

The thyroid concentrates iodide from the serum and oxidizes it at the apical membrane, attaching it to tyrosyl residues within thyroglobulin (Tg) to make diiodotyrosine and moniodotyrosine. Major players in this process

### \*\*\* KEYWORD \*\*\*

pendrin

### \*\*\* Another Name \*\*\*

[GenBank]

• PDS

### [LocusLink]

• SLC26A4

• solute carrier family

26, member 4

• PDS

• DFNB4

• SLC26A8

• solute carrier family

26, member 6

• DKFZ 58851422

### [BLAST]

• down-regulated in

adenoma

• down-regulated in

adenoma protein

• sulfate anion

transporter 1

• diastrophic

dysplasia

• sulfate anion

transporter

• solute carrier family

26

• high affinity sulfate

permease

• Prokallin sulfate

transport protein

• Yp0000p

【書類名】 要約書

【要約】

【課題】 遺伝子あるいはタンパク質が名前以外では検索できないなどの生物情報データベースの検索を行う際の問題を解決するような生物情報検索システム及び検索方法を提供する。

【解決手段】 ユーザから生物情報データベースへの問合せを受け取ったユーザ入力受付部 11 がユーザクエリ 111 を生成し、クエリ生成部 12 が生物情報データベースクエリ 121 を生成し、生物情報データベース検索部 13 が実際に生物情報データベース 999 の検索を行って、生物情報データベース検索結果 133 を出力し、検索結果解析部 14 が検索結果の解析を行って解析結果 142 を出力し、出力制御部 15 が、解析結果を受けて順次出力を行うとともに未検索分について再検索クエリ 151 を生成してクエリ生成部 12 に送り、出力情報指定保持部 16 が、検索結果の抽出及び出力に関する条件をあらかじめ保持しておく。

【選択図】 図 1



特願 2003-109054

出 願 人 履 歴 情 報

識別番号

[000004237]

1. 変更年月日

1990年 8月29日

[変更理由]

新規登録

住 所

東京都港区芝五丁目7番1号

氏 名

日本電気株式会社